I. P. NATANSON

CONSTRUCTIVE FUNCTION THEORY

VOLUME II

APPROXIMATION IN MEAN

A volume, equally elegant in argumentation and in presentation, of one of the most interesting mathematical branches, dealing with approximate representation of arbitrary functions.

Although the recent impressive development of Constructive Function Theory has made it a necessity for mathematicians and engineers alike, the lack of a systematic textbook has hitherto kept it from being a regular feature in the mathematics curriculum.

This work, based on lectures by the author at Leningrad University, is designed to fill the textbook gap:

VOLUME I deals chiefly with the approximation of arbitrary real continuous functions by algebraic polynomials, and the approximation of periodic functions by trigonometric polynomials, especially the Fourier series.

VOLUME II deals with the theory of approximation in mean. While we avail ourselves again of algebraic and trigonometric polynomials as approximation functions, the functions thus approximated are of a more general kind, and the criterion for the accuracy achieved is changed. This different view leads us to new problems and new results.

VOLUME III, applying the method of interpolation, again changes the criterion for the accuracy of approximation. It will not be a general maximal distance of corresponding points of the approximated and the approximating functions, but the factual coincidence in certain selected points called nodes of interpolation. All these angles of

continued on back flap

CONSTRUCTIVE FUNCTION THEORY

APPROXIMATION IN MEAN

I. P. NATANSON

CONSTRUCTIVE FUNCTION THEORY

Volume II

APPROXIMATION IN MEAN

Translated by

JOHN R. SCHULENBERGER

FREDERICK UNGAR PUBLISHING CO.

NEW YORK

Copyright © 1965 by
Frederick Ungar Publishing Co., Inc.

Printed in the United States of America

Library of Congress Catalog Card No. 64—15689

CONTENTS

Chapter I. THE SPACE $L^2_{p(x)}$

1. The Problem 7
2. The Weight Function. The Space $L^2_{p(x)}$ 9
3. Convergence in Mean 11
4. Classes of Functions Dense in $L^2_{p(x)}$ 15

Chapter II. ORTHOGONAL SYSTEMS

1. Orthogonality. Examples 19
2. The FOURIER Coefficients 23
3. Completeness and Closedness 29

Chapter III. LINEARLY INDEPENDENT SYSTEMS OF FUNCTIONS

1. Linear Independence. The GRAM Determinant. The SCHMIDT Theorem 32
2. Approximation by Linearly Independent Functions 36
3. The MÜNTZ Theorems 40

Chapter IV. GENERAL PROPERTIES OF ORTHOGONAL POLYNOMIALS

1. Foundations 45
2. Roots of Orthogonal Polynomials. A Recursion Formula 50
3. Relation to the Theory of Continued Fractions 58
4. The CHRISTOFFEL-DARBOUX Formula. Convergence of Orthogonal Expansions. 66
5. Changing the Weight Function 74

Chapter V. THE LEGENDRE POLYNOMIALS

1. The RODRIGUES Formula 82
2. The Generating Function 88
3. The LAPLACE Integral 91
4. Expansions in Terms of LEGENDRE Polynomials 94

Chapter VI. The Jacobi Polynomials

1. Generalization of the Rodrigues Formula 102
2. The Recursion Formula, Generating Function, and Differential Equation 107
3. Estimates of the Jacobi Polynomials. The Expansion Problem. 109
4. Tchebysheff Polynomials of the Second Kind 114
5. The Jacobi Polynomials for $\alpha = \frac{1}{2}$, $\beta = -\frac{1}{2}$ 121

Chapter VII. The Moment Problem for Finite Intervals

1. Formulation of the Problem 126
2. The Theorems of F. Hausdorff 130
3. Linear Functionals in C and L^2 136
4. Positive Definite Sequences 142

Chapter VIII. Infinite Intervals

1. Introduction 147
2. The Laguerre Polynomials 151
3. The Generalized Laguerre Polynomials 153
4. The Hermite Polynomials 156
5. The Moment Problem for Infinite Intervals 159
6. The Favard Theorem 167

Bibliography 171

Index 175

Index of Symbols 176

CHAPTER I

THE SPACE $L^2_{p(x)}$

§1. The Problem

In this part of the book, as in the first part, we shall be mainly concerned with the approximation of an arbitrary function by means of polynomials. However, a different criterion of accuracy will now be used.

Up until now we have considered the polynomial approximation of a function continuous in the interval $[a, b]$ to be better the smaller the quantity

$$\max_{a \leqq x \leqq b} |P(x) - f(x)|. \tag{1}$$

We will now call a polynomial $P(x)$ a good approximation to the function $f(x)$ if the value of the integral [1]

$$\int_a^b [P(x) - f(x)]^2 \, dx \tag{2}$$

is small.

It is to be noted that the requirement of continuity of $f(x)$ is a natural one when we use criterion of accuracy (1). For when we approximate a function $f(x)$ arbitrarily close by means of polynomials according to this method, this means that $f(x)$ is the limit function of a *uniformly* convergent sequence of polynomials and is hence itself continuous.

The matter is quite different when the approximation criterion involving the integral (2) is used. For example, suppose that in the interval $[-1, +1]$ the function $f(x)$ is given as follows:

$$f(x) = \begin{cases} 0 \text{ for } -1 \leqq x \leqq 0, \\ 1 \text{ for } \quad 0 < x \leqq 1. \end{cases}$$

Although this function is discontinuous at $x = 0$, there exist polynomials $P(x)$ for which

$$\int_{-1}^{+1} [P(x) - f(x)]^2 \, dx$$

becomes arbitrarily small.

[1] We shall ordinarily use somewhat more general integrals than (2) as a criterion of approximation of $P(x)$ to $f(x)$; however, in order not to complicate matters from the outset, we will in the beginning adopt the criterion given in the text.

In order to show this, we define a function $\varphi(x)$ as follows:

$$\varphi(x) = \begin{cases} 0 \text{ for } -1 \leqq x \leqq 0 \\ 1 \text{ for } \dfrac{1}{n} \leqq x \leqq 1, \end{cases}$$

and in the interval $\left[0, \frac{1}{n}\right]$ $\varphi(x)$ increases linearly from 0 to 1. It is obvious that $\varphi(x)$ is continuous in $[-1, +1]$, and moreover

$$\int_{-1}^{+1} [\varphi(x) - f(x)]^2 \, dx = \int_{0}^{1/n} [\varphi(x) - f(x)]^2 \, dx < \frac{1}{n},$$

since $|\varphi(x) - f(x)| \leqq 1$. On the basis of the WEIERSTRASS theorem, there exists a polynomial $P(x)$ which satisfies the inequality

$$|P(x) - \varphi(x)| < \frac{1}{\sqrt{2n}}$$

for all x in the interval $[-1, +1]$. For this polynomial it follows that

$$\int_{-1}^{+1} [P(x) - \varphi(x)]^2 \, dx < \frac{1}{n}.$$

From

$$(a + b)^2 \leqq 2(a^2 + b^2)$$

it follows that

$$[P(x) - f(x)]^2 \leqq 2\{[P(x) - \varphi(x)]^2 + [\varphi(x) - f(x)]^2\},$$

and therefore

$$\int_{-1}^{+1} [P(x) - f(x)]^2 \, dx < \frac{4}{n},$$

whence our assertion follows, since n may be chosen arbitrarily large.

Therefore, in our new approach the requirement of continuity of the function to be approximated is unnecessary. We shall therefore no longer impose this restriction, but rather admit discontinuous functions as well. However, these functions may by no means be entirely arbitrary; they must rather be compatible with the existence of the integral (2). If we were to adopt the RIEMANN definition of the integral, this would mean a restriction to those discontinuous functions whose points of discontinuity form a set of measure zero. Such a restriction however would not be intrinsic in the theory now to be developed, but would rather stem from the integral concept employed.

Therefore, in order to achieve the most complete presentation possible, we adopt not the RIEMANN concept of the integral, but rather the LEBESGUE concept. This, to be sure, requires that the reader be familiar with the foundations of the theory of functions of a real variable, which we shall presuppose in what follows.

§2. The Weight Function. The Space $L^2_{p(x)}$

Let a non-negative summable function $p(x)$ be given on the interval $[a, b]$. In view of the special role which this function will play in what follows, we will call it a *weight function* or briefly—a *weight*. We agree once and for all to consider only those weights $p(x)$ which assume the value zero only on a set of measure zero; we shall henceforth make no special mention of this stipulation.

Each weight function $p(x)$ determines two classes of measurable functions on $[a, b]$: the class $L_{p(x)}$ of those functions $f(x)$ for which the product $p(x)f(x)$ is summable and the class $L^2_{p(x)}$ of those functions $f(x)$ for which the product $p(x)f^2(x)$ is summable. If $p(x) = 1$, we designate these classes simply by L and L^2. If in designating these classes it is necessary to indicate the interval of integration as well, we write $L_{p(x)}([a, b])$, $L^2_{p(x)}([a, b])$, $L([a, b])$, and $L^2([a, b])$.

As is shown by the inequality

$$|f(x)| \leqq \frac{1 + f^2(x)}{2},$$

$L^2_{p(x)}$ is contained in $L_{p(x)}$.

The inequality

$$|f(x)g(x)| \leqq \frac{f^2(x) + g^2(x)}{2}$$

further shows that the product of two functions in $L^2_{p(x)}$ belongs to $L_{p(x)}$. This together with the identity

$$(f \pm g)^2 = f^2 \pm 2fg + g^2$$

implies that the sum and difference of two functions in $L^2_{p(x)}$ also belong to $L^2_{p(x)}$. Finally, it is fundamental that with $f(x)$ all functions $cf(x)$, where c is a constant, belong to $L^2_{p(x)}$.

Theorem 1. *If $f(x) \in L^2_{p(x)}$ and $g(x) \in L^2_{p(x)}$, then the following two inequalities, which we shall call the Schwarz inequalities, hold:*

$$\left[\int_a^b p(x)f(x)g(x)\,dx\right]^2 \leqq \left[\int_a^b p(x)f^2(x)\,dx\right]\left[\int_a^b p(x)g^2(x)\,dx\right] \tag{3}$$

$$\sqrt{\int_a^b p(x)[f(x)+g(x)]^2\,dx} \leqq \sqrt{\int_a^b p(x)f^2(x)\,dx} + \sqrt{\int_a^b p(x)g^2(x)\,dx}. \tag{4}$$

Proof. To prove inequality (3) we define

$$\psi(z) = \int_a^b p(x)[zf(x)+g(x)]^2\,dx = Az^2 + 2Bz + C,$$

where

$$A = \int_a^b p(x)f^2(x)\,dx, \quad B = \int_a^b p(x)f(x)g(x)\,dx, \quad C = \int_a^b p(x)g^2(x)\,dx.$$

If $A = 0$, then $f(x) = 0$ (as is usual in the measure theory of functions, we do not distinguish between two functions which differ only on a set of measure zero), and inequality (3) becomes the equality $0 = 0$. If $A > 0$, then (3) is obtained by observing that $\psi(z) \geqq 0$ and

$$\psi\left(-\frac{B}{A}\right) = \frac{AC - B^2}{A}.$$

This completes the proof of (3). We now write (3) in the form

$$\int_a^b pfg\,dx \leqq \sqrt{\int_a^b pf^2\,dx}\sqrt{\int_a^b pg^2\,dx},$$

multiply by 2, and add to both sides the expression

$$\int_a^b pf^2\,dx + \int_a^b pg^2\,dx;$$

we hereby obtain the inequality

$$\int_a^b p(f+g)^2\,dx \leqq \left[\sqrt{\int_a^b pf^2dx} + \sqrt{\int_a^b pg^2\,dx}\right]^2,$$

which is equivalent to (4).

With each function $f(x) \in L^2_{p(x)}$ we now associate the number

$$\|f\| = \sqrt{\int_a^b p(x)f^2(x)\,dx},$$

which we call the *norm* of the function $f(x)$; the norm has properties similar to those of the absolute value of a number:

I. $\|f\| \geqq 0$, *and* $\|f\| = 0$ *if and only if* $f = 0$.

II. $\|cf\| = |c| \cdot \|f\|$, *and in particular* $\|-f\| = \|f\|$.

III. $\|f+g\| \leqq \|f\| + \|g\|$.

This norm concept makes it possible for us to introduce a useful geometric interpretation.

Let E be a set of elements $x, y, z, \ldots$ of any type whatsoever.

With each pair of elements x, y let a real number $r(x, y)$ be associated with the properties

1. $r(x, y) \geqq 0$, *and* $r(x, y) = 0$ *if and only if* $x = y$.
2. $r(x, y) = r(y, x)$.
3. $r(x, z) \leqq r(x, y) + r(y, z)$.

When all these conditions are met, the set E is said to form a *metric space, and* $r(x, y)$ *is called the distance between* x *and* y.

If we define the distance between two functions f and g in $L^2_{p(x)}$ as

$$r(f, g) = \|f - g\|,$$

then $L^2_{p(x)}$ becomes a metric space.

§3. Convergence in Mean

Definition 1. An element f of the space $L^2_{p(x)}$ is called a *limit* of a sequence $f_1, f_2, f_3, \ldots, f_n, \ldots$ of elements of the same space if

$$\lim_{n\to\infty} \|f_n - f\| = 0.$$

We will usually abbreviate this relation

$$\lim_{n\to\infty} f_n = f, f_n \to f,$$

although its function-theoretic meaning is actually

$$\lim_{n\to\infty} \int_a^b p(x)[f_n(x) - f(x)]^2 dx = 0.$$

Convergence of this type we call *convergence in mean with the weight function* $p(x)$.

Theorem 1. *A sequence of elements of* $L^2_{p(x)}$ *cannot have two distinct limits.*

Proof. Suppose $f_n \to f$ and $f_n \to g$. Passing to the limit in the inequality

$$0 \leqq \|f - g\| \leqq \|f - f_n\| + \|f_n - g\|$$

gives the relation $\|f - g\| = 0$, and hence $f = g$.

Theorem 2. *If a sequence of functions* $\{f_n(x)\}$ *converges in mean to the function* $f(x)$, *then it is possible to select from it a subsequence* $\{f_{n_k}\}$ *which converges to* $f(x)$ *almost everywhere.*

The proof of this theorem is based on the following important theorem from function theory (which we do not prove here):

Theorem (B. LEVI).[2] *Given a sequence of non-negative measurable functions*

$$u_1(x), \quad u_2(x), \quad u_3(x), \quad \ldots$$

on the interval $[a, b]$ *such that*

$$\sum_{k=1}^{\infty} \int_a^b u_k(x)\, dx < +\infty,$$

then

$$\lim_{n\to\infty} u_n(x) = 0$$

almost everywhere in $[a, b]$.

Proof of Theorem 2. Choose a sequence of indices $n_1 < n_2 < n_3 < \cdots$ such that

$$\int_a^b p(x)[f_{n_k}(x) - f(x)]^2 dx < \frac{1}{k!}.$$

Then from the LEVI theorem

$$\lim_{k\to\infty} p(x)[f_{n_k}(x) - f(x)]^2 = 0$$

almost everywhere in $[a, b]$, and since $p(x)$ is positive almost everywhere, the proof of the theorem is completed.

Theorem 3. *If* $f_n \to f$, *then* $\|f_n\| \to \|f\|$.

Proof. From

$$\|f\| \leqq \|f_n\| + \|f - f_n\|, \quad \|f_n\| \leqq \|f\| + \|f_n - f\|$$

one obtains

$$|\, \|f_n\| - \|f\|\, | \leqq \|f_n - f\|,$$

whence the assertion follows.

Definition 2. A sequence $\{f_n\} \in L^2_{p(x)}$ is called *convergent in itself* if to every $\varepsilon > 0$ there corresponds an integer N such that $n > N$ and $m > N$ imply

$$\|f_n - f_m\| < \varepsilon. \tag{5}$$

[2] B. LEVI stated the theorem in a more general context. Cf. I. P. Natanson [5], p. 141.

Theorem 4. *If a sequence of functions has a limit, then it is also convergent in itself.*

Proof. Suppose $f_n \to f$. Then given any $\varepsilon > 0$, there exists an integer N such that for all $n > N$

$$\|f_n - f\| < \frac{\varepsilon}{2}.$$

Choosing $n > N$ and $m > N$,

$$\|f_n - f_m\| \leqq \|f_n - f\| + \|f - f_m\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The converse is also true:

Theorem 5 (E. FISCHER [1]). *If a sequence of functions converges in itself, then it has a limit.*

This theorem expresses a property of the space $L^2_{p(x)}$ which is called *completeness.*

Proof. Choose n_k such that for all $n \geqq n_k$ and $m \geqq n_k$

$$\|f_n - f_m\| < \frac{1}{k!};$$

we may assume that $n_1 < n_2 < n_3 < \cdots$. Then in particular

$$\|f_{n_{k+1}} - f_{n_k}\| < \frac{1}{k!},$$

so that the series

$$\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|$$

converges.

Applying inequality (3) to the two functions

$$f = |f_{n_{k+1}} - f_{n_k}| \quad \text{and} \quad g = 1,$$

we obtain

$$\int_a^b p(x)|f_{n_{k+1}}(x) - f_{n_k}(x)|dx \leqq \sqrt{\int_a^b p(x)\,dx}\,\|f_{n_{k+1}} - f_{n_k}\|.$$

Therefore, the series

$$\sum_{k=1}^{\infty} \int_a^b p(x)|f_{n_{k+1}}(x) - f_{n_k}(x)|dx$$

also converges, whence from the LEVI theorem it follows that the series

$$\sum_{k=1}^{\infty} p(x)|f_{n_{k+1}}(x) - f_{n_k}(x)|$$

converges almost everywhere and thus also the series

$$f_{n_1}(x) + \sum_{k=1}^{\infty} \{f_{n_{k+1}}(x) - f_{n_k}(x)\}.$$

Convergence of the last series is equivalent to the existence of a finite limit

$$\lim_{k\to\infty} f_{n_k}(x).$$

We now define a function $f(x)$ which is equal to this limit at all points where the limit exists and is zero at all other points; we then show that this function belongs to $L^2_{p(x)}$ and is the limit function of $f_n(x)$. Let $\varepsilon > 0$ be given. Then there exists an integer N such that for $n > N$ and $m > N$ the inequality (5) is satisfied. Fix $n > N$. For sufficiently large values of k, $n_k > N$, so that

$$\|f_n - f_{n_k}\| < \varepsilon.$$

At this point we need still another theorem from function theory [3] (which we again do not prove here):

Theorem (P. FATOU). *If a sequence of non-negative, measurable functions* $\varphi_1(x), \varphi_2(x), \ldots$ *converges to a function* $\psi(x)$ *almost everywhere on the interval* $[a, b]$, *and if for all* k

$$\int_a^b \varphi_k(x)\,dx \leqq A,$$

then

$$\int_a^b \psi(x)\,dx \leqq A.$$

In our case $\varphi_k(x)$ is the function $p(x)[f_n(x) - f_{n_k}(x)]^2$ (where n is fixed), $\psi(x)$ is the function $p(x)[f_n(x) - f(x)]^2$, and A is the quantity ε^2. This gives

$$\int_a^b p(x)[f_n(x) - f(x)]^2\,dx \leqq \varepsilon^2. \tag{6}$$

It follows from this that the difference $f_n(x) - f(x)$, and therefore the function $f(x)$ itself belongs to $L^2_{p(x)}$. Since only the condition $n > N$ was necessary for the validity of inequality (6), the theorem is therewith proved.

[3] See, for example, I. P. NATANSON [5], p. 140.

In addition to convergence in mean, we shall also be concerned with so-called "weak convergence." A sequence $\{f_n(x)\} \in L^2_{p(x)}$ is said to converge weakly to $f(x) \in L^2_{p(x)}$ if for each function $g(x) \in L^2_{p(x)}$ the following relation holds:

$$\lim_{n\to\infty} \int_a^b p(x) f_n(x) g(x)\, dx = \int_a^b p(x) f(x) g(x)\, dx.$$

From the SCHWARZ inequality

$$\left| \int_a^b p f_n g\, dx - \int_a^b p f g\, dx \right| \leqq \sqrt{\int_a^b p g^2\, dx} \sqrt{\int_a^b p (f_n - f)^2\, dx};$$

therefore convergence of a sequence in mean implies weak convergence (to the same limit function).

§ 4. Classes of Functions Dense in $L^2_{p(x)}$

Suppose that E is a metric space, and A is a subset of this space. If each element of E can be represented as the limit of a sequence of elements of A, then A is said to be *everywhere dense* in E. A necessary and sufficient condition for this is clearly that for any given $\varepsilon > 0$ it is possible to find for each $\in E$ an element $g \in A$ such that the condition

$$r(f, g) < \varepsilon$$

is satisfied.

Since we have assumed the weight function $p(x)$ to be summable at the beginning, all bounded, measurable functions, and a fortiori all continuous and all "step functions" [4], belong to $L^2_{p(x)}$.

We now adopt the following notation: M shall designate the class of all bounded, measurable functions, C the class of all continuous functions, S the class of all step functions (we of course always refer to functions defined on the interval $[a, b]$), P the class of all polynomials, and T the class of all trigonometric polynomials.

Theorem. *Each of the classes M, C, S, and P is everywhere dense in $L^2_{p(x)}$. If the interval of definition $[a, b]$ is of length 2π, then T is also dense in $L^2_{p(x)}$.*

Proof. 1. Let $f(x) \in L^2_{p(x)}$. Given $\varepsilon > 0$, it is possible to find a $\delta > 0$ such that for each measurable set e in $[a, b]$ of measure $me < \delta$ the inequality

$$\int_e p(x) f^2(x)\, dx < \varepsilon^2$$

[4] A function is called a step function (or "stepped") in $[a, b]$, if there exists a finite number of points $a = c_1 < c_2 < \cdots < c_n = b$ such that the function is constant in each interval (c_i, c_{i+1}).

holds; that it is possible to find such a δ follows from the absolute continuity of the integral.

Since $f(x) \in L^2_{p(x)}$ and is hence finite almost everywhere, it follows that

$$m \prod_{n=1}^{\infty} E(|f| > n) = 0. \tag{7}$$

On the other hand

$$E(|f| > 1) \supset E(|f| > 2) \supset \cdots \tag{8}$$

and from (7) and (8) it now follows that

$$\lim_{n \to \infty} mE(|f| > n) = 0.$$

Therefore, a number n can be given for which

$$mE(|f| > n) < \delta.$$

Fix this n and put

$$g(x) = \begin{cases} f(x), & \text{if } |f(x)| \leqq n, \\ 0, & \text{if } |f(x)| > n. \end{cases}$$

It is clear that $g(x) \in M$ and

$$\|f - g\|^2 = \int_a^b p(f-g)^2\,dx = \int_{E(f \neq g)} p(f-g)^2\,dx = \int_{E(|f|>n)} pf^2\,dx < \varepsilon^2.$$

This proves the theorem for M.

2. Again let $f(x) \in L^2_{p(x)}$ and $\varepsilon > 0$. From what was proved above, we can find a $g(x) \in M$ such that $\|f - g\| < \frac{\varepsilon}{2}$. Let $|g(x)| \leqq K$. From a well-known theorem due to N. N. LUSIN, there exists a continuous function $\varphi_\delta(x)$ which satisfies both the conditions

$$mE(g \neq \varphi_\delta) < \delta, \quad |\varphi_\delta(x)| \leqq K,$$

where $\delta > 0$ may be given arbitrarily. For this function φ_δ we obtain

$$\|g - \varphi_\delta\|^2 = \int_{E(g \neq \varphi_\delta)} p(g - \varphi_\delta)^2\,dx \leqq 4K^2 \int_{E(g \neq \varphi_\delta)} p(x)\,dx$$

By virtue of the absolute continuity of the integral, δ may be chosen so small that the right-hand side of the last inequality becomes less than $\frac{\varepsilon^2}{4}$. The following estimate is then valid for the function $\varphi_\delta(x) \in C$.

$$\|f - \varphi_\delta\| \leqq \|f - g\| + \|g - \varphi_\delta\| < \varepsilon.$$

3. Again let $f(x) \in L^2_{p(x)}$ and $\varepsilon > 0$. From what was just proved, there exists a function $\varphi \in C$ which satisfies the condition $\|f - \varphi\| < \frac{\varepsilon}{2}$.

Because of the uniform continuity of $\varphi(x)$ in $[a, b]$, we can partition this interval by means of a sequence of points $c_0 = a < c_0 < \cdots < c_n = b$ into subintervals $[c_i, c_{i+1}]$ such that the oscillation of $\varphi(x)$ in each of these subintervals becomes less than any prescribed number $\delta > 0$. We now define a step function $h(x)$ as follows: $h(b) = \varphi(b)$ and $h(x) = \varphi(c_i)$ for $c_i \leqq x < c_{i+1}$ $(i = 0, 1, \ldots, n - 1)$. For all $x \in [a, b]$, this step function satisfies the inequality

$$|h(x) - \varphi(x)| < \delta.$$

Therefore

$$\|h - \varphi\|^2 = \int_a^b p(x)[h(x) - \varphi(x)]^2\, dx \leqq \delta^2 \int_a^b p(x)\, dx.$$

By choosing δ sufficiently small the right-hand side of this inequality can be made less than $\frac{\varepsilon^2}{4}$ and hence $\|f - h\| < \varepsilon$. The theorem is thus proved for the class S.

4. For the class P the proof goes similarly, since from the first WEIERSTRASS theorem for any given $\delta > 0$ we can always find a polynomial $P(x)$ which satisfies the condition

$$|P(x) - \varphi(x)| < \delta.$$

5. Finally, if $b - a = 2\pi$ we first construct a continuous function $\varphi(x)$ such that $\|f - \varphi\| < \frac{\varepsilon}{2}$; we then introduce another continuous function $\psi(x)$ by defining

$$\psi(x) = \varphi(x) \text{ in } [a, b - \delta],\ \psi(b) = \varphi(a),$$

and $\psi(x)$ shall be linear in the interval $[b - \delta, b]$. δ is hereby any positive number, the choice of which we subsequently determine more precisely. If K is the maximum of $|\varphi(x)|$, then $|\psi(x)| \leqq K$ also. Therefore

$$\|\varphi - \psi\|^2 = \int_{b-\delta}^b p(\varphi - \psi)^2\, dx \leqq 4K^2 \int_{b-\delta}^b p(x)\, dx.$$

We can now choose δ so small that the right-hand side of this inequality becomes less than $\frac{\varepsilon^2}{16}$. Then

$$\|f - \psi\| < \tfrac{3}{4}\varepsilon.$$

At the same time $\psi(x)$ is also 2π-periodic, since $\psi(a) = \psi(b)$. From the second WEIERSTRASS theorem there then exists a trigonometric polynomial which approximates $\psi(x)$ arbitrarily close; the proof now proceeds as in the previous cases.

CHAPTER II

ORTHOGONAL SYSTEMS

§ 1. Orthogonality. Examples

Let $p(x)$ be a weight function defined on the interval $[a, b]$. Two functions $f(x)$ and $g(x)$ which satisfy the relation

$$\int_a^b p(x)f(x)g(x)\,dx = 0,$$

are said to be *orthogonal on the interval* $[a, b]$ *with the weight function* $p(x)$. If $p(x) \equiv 1$, the weight is not explicitly stated, and one says simply that $f(x)$ and $g(x)$ *are orthogonal on the interval* $[a, b]$.

If all the functions in a (finite or infinite) system of functions

$$\omega_1(x),\ \omega_2(x),\ \omega_3(x),\ \ldots \qquad (a \leqq x \leqq b) \quad (9)$$

are pairwise orthogonal on $[a, b]$ with the weight function $p(x)$, we call the system an *orthogonal system of functions with the weight function* $p(x)$. If $p(x) \equiv 1$, we do not mention the weight. Examples are the trigonometric system

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cos 3x,\ \ldots \qquad (10)$$

which is orthogonal on $[-\pi, \pi]$ and the system of TCHEBYSHEFF polynomials

$$T_0(x),\ T_1(x),\ T_2(x),\ \ldots \qquad (11)$$
$$(T_n(x) = \cos\,(n \arccos x)),$$

which is orthogonal on $[-1, +1]$ with the weight function $(1 - x^2)^{-1/2}$.

From the beginning we confine ourselves to orthogonal systems in which first of all none of the functions $\omega_k(x)$ is equivalent to zero, and secondly each belongs to the class $L^2_{p(x)}$. The integrals

$$A_k = \int_a^b p(x)\omega_k^2(x)\,dx \qquad (12)$$

are therefore finite positive numbers $0 < A_k < +\infty$.

Those orthogonal systems in which $A_k = 1$ for each index k are especially useful. Such systems are called *normalized orthogonal systems* or briefly—*orthonormal systems.*

It is easily seen that from the orthogonality of the system (9) it follows that the system

$$\frac{\omega_1(x)}{\sqrt{A_1}}, \quad \frac{\omega_2(x)}{\sqrt{A_2}}, \quad \frac{\omega_3(x)}{\sqrt{A_3}}, \quad \ldots \tag{13}$$

is orthonormal. The transition from (9) to (13) is called *normalization* of the functions of the system.

In addition to (10) and (11), we now give several other examples of orthogonal systems.

1°. Each of the two systems

$$1, \cos x, \cos 2x, \cos 3x, \ldots, \tag{14}$$

$$\sin x, \sin 2x, \sin 3x, \ldots \tag{15}$$

is orthogonal on $[0, \pi]$.

2°. STURM-LIOUVILLE systems. We consider the differential equation

$$\frac{d^2y}{dx^2} + \lambda p(x)y = 0, \tag{16}$$

in which $p(x) > 0$ is a continuous function defined on the interval $[a, b]$ and λ is a numerical parameter.

We are here interested in those continuous solutions $y = y(x)$ of this equation which satisfy the boundary conditions

$$y(a) = y(b) = 0. \tag{17}$$

One such solution is the trivial solution $y = 0$. Every other solution which is not identically zero is called an *eigenfunction* of our equation. Such eigenfunctions do not exist for every value of the parameter λ. It is therefore convenient to single out those values of λ for which eigenfunctions do exist and give them a particular name: they are usually called *eigenvalues* of the equation.

In the theory of differential equations it is proved that such eigenvalues always exist and that to each of these eigenvalues there corresponds exactly one (up to a constant factor) well-defined eigenfunction. The eigenvalues of the equation may be written in a row with the corresponding eigenfunctions directly below them:

λ_1	λ_2	λ_3	...
$y_1(x)$	$y_2(x)$	$y_3(x)$	

$(\lambda_i \neq \lambda_k)$.

Theorem 1. *The system of eigenfunctions*

$$y_1(x),\ y_2(x),\ y_3(x),\ \ldots \tag{18}$$

is orthogonal on $[a, b]$ *with weight function* $p(x)$.

Proof. Let $i \neq k$. Then

$$\frac{d^2y_i}{dx^2} + \lambda_i p(x) y_i = 0, \qquad \frac{d^2y_k}{dx^2} + \lambda_k p(x) y_k = 0.$$

Multiply the first of these two equations by y_k, the second by y_i, and then subtract the second from the first. Using the identity

$$y_k y_i'' - y_i y_k'' = \frac{d}{dx}(y_k y_i' - y_i y_k'),$$

we obtain

$$\frac{d}{dx}(y_k y_i' - y_i y_k') + (\lambda_i - \lambda_k) p(x) y_i y_k = 0;$$

whence it follows that

$$[y_k y_i' - y_i y_k']_a^b + (\lambda_i - \lambda_k) \int_a^b p(x) y_i(x) y_k(x)\, dx = 0.$$

Now since $y_i(x)$ and $y_k(x)$ satisfy condition (17),

$$[y_k y_i' - y_i y_k']_a^b = 0,$$

and therefore since $\lambda_i \neq \lambda_k$

$$\int_a^b p(x) y_i(x) y_k(x)\, dx = 0,$$

which completes the proof.

A system of the type (18) is called a STURM-LIOUVILLE *system*. The system (15) which corresponds to the differential equation

$$\frac{d^2y}{dx^2} + \lambda y = 0 \qquad (y(0) = y(\pi) = 0)$$

with eigenvalues

$$\lambda_k = k^2 \qquad (k = 1, 2, 3, \ldots)$$

is an example of such a system.

3°. The RADEMACHER system [1]. We partition the interval $[0, 1]$ by means of the points

$$0,\ \frac{1}{2^k},\ \frac{2}{2^k},\ \ldots,\ \frac{n}{2^k},\ \frac{n+1}{2^k},\ \ldots,\ \frac{2^k - 1}{2^k},\ 1 \tag{19}$$

into 2^k parts and define a function $r_k(x)$ as follows:

$$r_k\left(\frac{n}{2^k}\right) = 0 \qquad (n = 0, 1, 2, \ldots, 2^k),$$

$$r_k(x) = (-1)^n \text{ for } \frac{n}{2^k} < x < \frac{n+1}{2^k} \qquad (n = 0, 1, 2, \ldots, 2^k - 1).$$

The system of functions

$$r_1(x),\ r_2(x),\ r_3(x),\ \ldots \tag{20}$$

is called the RADEMACHER *system.*

Theorem 2. *The* RADEMACHER *system is orthonormal on the interval* [0, 1].

Proof. Since $r_k^2(x) = 1$ (except at the points (19)),

$$r_k^2(x) = 1$$

$$\int_0^1 r_k^2(x)\,dx = 1. \tag{21}$$

Now let k and i be two natural numbers such that $k < i$. Then

$$\int_0^1 r_k(x)r_i(x)\,dx = \sum_{n=0}^{2^k-1} \int_{\frac{n}{2^k}}^{\frac{n+1}{2^k}} r_k(x)r_i(x)\,dx = \sum_{n=0}^{2^k-1} (-1)^n \int_{\frac{n}{2^k}}^{\frac{n+1}{2^k}} r_i(x)\,dx.$$

By the points $\frac{m}{2^i}$ the interval $\left[\frac{n}{2^k}, \frac{n+1}{2^k}\right]$ is partitioned into an even number of subintervals:

$$\left[\frac{n\cdot 2^{i-k}}{2^i}, \frac{n\cdot 2^{i-k}+1}{2^i}\right],$$

$$\left[\frac{n\cdot 2^{i-k}+1}{2^i}, \frac{n\cdot 2^{i-k}+2}{2^i}\right], \ \ldots, \ \left[\frac{n\cdot 2^{i-k}+2^{i-k}-1}{2^i}, \frac{(n+1)\cdot 2^{i-k}}{2^i}\right],$$

in which $r_i(x)$ is alternately $+1$ and -1; therefore

$$\int_{\frac{n}{2^k}}^{\frac{n+1}{2^k}} r_i(x)\,dx = 0;$$

whence

$$\int_0^1 r_k(x)r_i(x)\,dx = 0. \tag{22}$$

With (21) and (22) the proof of the theorem is completed.

$$\text{Putting sign } z = \begin{cases} +1 \text{ for } z > 0, \\ 0 \text{ for } z = 0, \\ -1 \text{ for } z < 0, \end{cases}$$

the function $r_k(x)$ can be written in the form

$$r_k(x) = \text{sign } [\sin (2^k \pi x)].$$

Indeed, $\sin (2^k \pi x) = 0$ for $x = \frac{n}{2^k}$; however, if $\frac{n}{2^k} < x < \frac{n+1}{2^k}$, then $n\pi < 2^k \pi x < (n+1)\pi$ so that $\sin (2^k \pi x)$ is positive or negative depending on whether n is even or odd.

§ 2. The FOURIER Coefficients

Suppose that an orthogonal system (9) with weight function $p(x)$ is given on the interval $[a, b]$. Let us consider a linear combination of a finite number of functions of this system:

$$f(x) = c_1\omega_1(x) + c_2\omega_2(x) + \cdots + c_n\omega_n(x). \tag{23}$$

Multiplying (23) by $p(x)\omega_k(x)$ $(k = 1, 2, \ldots, n)$ and integrating over $[a, b]$, we obtain by (12):

$$\int_a^b p(x)f(x)\omega_k(x)\,dx = A_k c_k,$$

whence

$$c_k = \frac{1}{A_k}\int_a^b p(x)f(x)\omega_k(x)\,dx. \tag{24}$$

The coefficients c_k in equation (23) are therefore uniquely determined by $f(x)$.

In the special case of the trigonometric system (10) the formulas (24) are the well-known formulas for the FOURIER coefficients of $f(x)$. For this reason, one calls the numbers (24) FOURIER *coefficients of* $f(x)$ *in the system* (9) in the general case as well. To compute these coefficients it is by no means necessary that $f(x)$ be a linear combination of the functions $\omega_k(x)$, but rather only that the integrals appearing in (24) exist. This is the case, as we know, whenever $f(x) \in L^2_{p(x)}$.

For each function $f(x) \in L^2_{p(x)}$ one can determine the FOURIER coefficients c_k and form the series

$$\sum_{k=1}^{\infty} c_k\omega_k(x) \tag{25}$$

(the system (9) is hereby assumed to be infinite). This series is called the FOURIER *series of the function* $f(x)$ *in the system* (9). Of course, nothing has yet been said about the convergence of this series; on the contrary, from the first part of the book we are familiar with an example of a continuous function whose trigonometric FOURIER series is not everywhere convergent. Hence, at the moment there is no justification for writing the equation

$$f(x) = \sum_{k=1}^{\infty} c_k \omega_k(x).$$

However, in order to indicate the relation of the series (25) to the function $f(x)$ one usually writes

$$f(x) \sim \sum_{k=1}^{\infty} c_k \omega_k(x).$$

We now turn to a problem which will show us the concept of FOURIER coefficients in an entirely new light.

We go back again to the system (9) and ask the question: what coefficients $a_1, a_2, \ldots, a_n$ must one choose in order that the linear combination

$$U(x) = \sum_{k=1}^{n} a_k \omega_k(x) \tag{26}$$

(for fixed n) approximate a particular function $f(x)$ in $L^2_{p(x)}$ with the smallest possible mean error, i.e. such that the integral

$$\|U - f\|^2 = \int_a^b p(x)[U(x) - f(x)]^2 \, dx \tag{27}$$

be a minimum.

It turns out that the solution of this problem is extremely simple. For on the one hand

$$\int_a^b p(x) f(x) U(x) \, dx = \sum_{k=1}^{n} a_k \int_a^b p(x) f(x) \omega_k(x) \, dx = \sum_{k=1}^{n} A_k a_k c_k,$$

where c_k are the coefficients of the FOURIER series of $f(x)$; on the other hand, from the orthogonality of the system (9)

$$\int_a^b p(x) U^2(x) \, dx = \sum_{k=1}^{n} A_k a_k^2,$$

so that we obtain

$$\|U - f\|^2 = \int_a^b p(x) f^2(x) \, dx - 2 \sum_{k=1}^{n} A_k a_k c_k + \sum_{k=1}^{n} A_k a_k^2$$

and from this

$$\|U - f\|^2 = \int_a^b p(x)f^2(x)\,dx - \sum_{k=1}^{n} A_k c_k^2 + \sum_{k=1}^{n} A_k(a_k - c_k)^2.$$

On the right-hand side of this equation only the sum

$$\sum_{k=1}^{n} A_k(a_k - c_k)^2$$

depends on the coefficients a_k; it assumes its smallest value (namely zero) when and only when

$$a_k = c_k \qquad (k = 1, 2, \ldots, n),$$

i.e. when the coefficients in question are chosen equal to the FOURIER coefficients of the function $f(x)$.
We have now proved

Theorem 1. (A. TOEPLER [1]). *Of all linear combinations*

$$U(x) = \sum_{k=1}^{n} a_k \omega_k(x)$$

the minimum of the integral

$$\|U - f\|^2 = \int_a^b p(x)[U(x) - f(x)]^2\,dx$$

is given by the partial sum of the FOURIER *series of* $f(x)$:

$$S_n(x) = \sum_{k=1}^{n} c_k \omega_k(x),$$

where the minimum has the value

$$\|S_n - f\|^2 = \int_a^b p(x)f^2(x)\,dx - \sum_{k=1}^{n} A_k c_k^2. \tag{28}$$

Since $\|S_n - f\|^2 \geqq 0$, (28) implies the BESSEL *inequality*

$$\sum_{k=1}^{n} A_k c_k^2 \leqq \int_a^b p(x)f^2(x)\,dx.$$

Now n was arbitrary (again assuming the system (9) to be infinite); the series

$$\sum_{k=1}^{\infty} A_k c_k^2$$

therefore converges, and the inequality

$$\sum_{k=1}^{\infty} A_k c_k^2 \leqq \int_a^b p(x) f^2(x)\, dx$$

holds, which (as already mentioned) is called the BESSEL inequality. If, in particular,

$$\sum_{k=1}^{\infty} A_k c_k^2 = \int_a^b p(x) f^2(x)\, dx, \tag{29}$$

then this is called the PARSEVAL *equality*. From (28) it follows that the PARSEVAL equality is equivalent to the relation

$$\lim_{n \to \infty} \|S_n - f\| = 0. \tag{30}$$

Theorem 2. *If $f(x)$ is a linear combination of a finite number of functions of the system, then it satisfies the* PARSEVAL *equality.*

Proof. As we have already seen, the coefficients c_k in the equation

$$f(x) = c_1\omega_1(x) + \cdots + c_n\omega_n(x) \tag{31}$$

are necessarily the FOURIER coefficients of the function $f(x)$. Therefore, multiplying (31) by $p(x)f(x)$ and integrating we obtain the PARSEVAL equality directly

$$\int_a^b p(x) f^2(x)\, dx = \sum_{k=0}^{\infty} A_k c_k^2.$$

The theorem can be proved in still another manner. We note that for $f(x) = \omega_k(x)$ the PARSEVAL equality is intrinsically fulfilled (it is contained in the definition of the numbers A_k), and we then prove

Theorem 3. *If the* PARSEVAL *equality holds for the functions*

$$f_1(x),\ f_2(x),\ \ldots,\ f_m(x),$$

then it also holds for every linear combination

$$f(x) = \sum_{i=1}^{m} a_i f_i(x). \tag{32}$$

Proof. We denote the nth partial sum of the FOURIER series of $f(x)$

$$\sum_{k=1}^{n} c_k \omega_k(x) \tag{33}$$

by $S_n[f]$; this notation is intended to stress the dependency of the partial sum (33) on the function $f(x)$. As is easily seen,

$$S_n[f] = \sum_{i=1}^{m} a_i S_n[f_i],$$

hence

$$f(x) - S_n[f] = \sum_{i=1}^{m} a_i \{f_i(x) - S_n[f_i]\}$$

and therefore

$$\|f - S_n[f]\| \leqq \sum_{i=1}^{m} |a_i| \cdot \|f_i - S_n[f_i]\|.$$

The PARSEVAL equality holds for each function $f_i(x)$ by hypothesis, so that the right-hand side of the last inequality tends to zero with increasing n; the left-hand side therefore satisfies the condition (30) which is equivalent to the validity of the PARSEVAL equality for $f(x)$.

Theorem 4. *A function $f(x)$ satisfies the* PARSEVAL *equality if and only if this function can be represented to arbitrary accuracy in the metric space $L^2_{p(x)}$ by linear combinations of the functions* (9). *That is to say: given $\varepsilon > 0$, there exists a linear combination*

$$U(x) = \sum_{k=1}^{n} a_k \omega_k(x)$$

which satisfies the condition

$$\|U - f\| < \varepsilon.$$

Proof. Necessity is obvious; as the linear combination $U(x)$ we may choose simply the nth partial FOURIER sum $S_n[f]$ with sufficiently large index n.

Conversely, suppose for every $\varepsilon > 0$ there exists such a linear combination $U(x)$. Then from the TOEPLER theorem

$$\|S_n - f\| \leqq \|U - f\|,$$

and therefore

$$\|S_n - f\| < \varepsilon.$$

We now turn to the following question: what conditions must a sequence of numbers $\{c_k\}$ satisfy in order to be a sequence of FOURIER coefficients for some function of $L^2_{p(x)}$? The answer to this is given by

Theorem 5. (F. RIESZ [1]—E. FISCHER [1]). *Let $\{\omega_k(x)\}$ be an orthogonal system on the interval $[a, b]$ with weight function $p(x)$, and let $\{c_k\}$ be a sequence of numbers such that the series*

$$\sum_{k=1}^{\infty} A_k c_k^2 \tag{34}$$

converges. Then in $L^2_{p(x)}$ there exists exactly one function $f(x)$ for which

1. *the numbers* c_k *are identical with the* FOURIER *coefficients of* $f(x)$ *in the system* $\{\omega_k(x)\}$;
2. *the* PARSEVAL *equality holds.*

Proof. We put

$$S_n(x) = \sum_{k=1}^{n} c_k \omega_k(x)$$

and convince ourselves that the sequence $\{S_n(x)\}$ converges in itself. To this end we compute $\|S_m - S_n\|$ where $m > n$:

$$\|S_m - S_n\|^2 = \left\| \sum_{k=n+1}^{m} c_k \omega_k \right\|^2 = \int_a^b p(x) \left[\sum_{k=n+1}^{m} c_k \omega_k(x) \right]^2 dx = \sum_{k=n+1}^{m} A_k c_k^2.$$

Since the series (34) converges, for given $\varepsilon > 0$ there exists a number N such that for $m > n > N$

$$\sum_{k=n+1}^{m} A_k c_k^2 < \varepsilon^2,$$

or what amounts to the same thing $\|S_m - S_n\| < \varepsilon$; but this means that the sequence $\{S_n\}$ converges in itself. By the completeness of the space $L^2_{p(x)}$ there exists a function $f(x) \in L^2_{p(x)}$ for which

$$\lim_{n \to \infty} \|S_n - f\| = 0. \tag{35}$$

This is the function sought. For the weak convergence of the sequence $\{S_n(x)\}$ to $f(x)$ follows from (35); therefore, for every function $g(x) \in L^2_{p(x)}$

$$\lim_{n \to \infty} \int_a^b p(x) S_n(x) g(x)\, dx = \int_a^b p(x) f(x) g(x)\, dx,$$

and for $g(x) = \omega_i(x)$ we have in particular

$$\int_a^b p(x) f(x) \omega_i(x)\, dx = \lim_{n \to \infty} \int_a^b p(x) S_n(x) \omega_i(x) dx.$$

If here $n \geqq i$, then

$$\int_a^b p(x) S_n(x) \omega_i(x)\, dx = \int_a^b p(x) \left[\sum_{k=1}^{n} c_k \omega_k(x) \right] \omega_i(x)\, dx = A_i c_i,$$

and hence

$$c_i = \frac{1}{A_i} \int_a^b p(x) f(x) \omega_i(x)\, dx.$$

The numbers c_k are therefore indeed the FOURIER coefficients of the function $f(x)$. But then $S_n(x)$ is the partial FOURIER sum of this function, and equation (35) asserts moreover that the PARSEVAL equality holds for $f(x)$.

We have still to show the uniqueness of this function. If there were two such functions, then according to condition 1 their FOURIER series would coincide. But since the PARSEVAL equality holds for both functions, this FOURIER series converges in mean to both functions, which is clearly possible only if the two functions coincide. This completes the proof of the theorem.

§3. Completeness and Closedness

Definition. An orthogonal system is called *closed* if for any arbitrary function of $L^2_{p(x)}$ the PARSEVAL equality holds.

From what has previously been said it follows that in the case of a closed orthogonal system the set of all linear combinations of this system is everywhere dense in $L^2_{p(x)}$.

Theorem 1. *If an orthogonal system is closed then for any two functions $f(x)$ and $g(x)$ of $L^2_{p(x)}$ the inequality*

$$\int_a^b p(x)f(x)g(x)\,dx = \sum_{k=1}^{\infty} A_k a_k b_k \tag{36}$$

holds, where a_k and b_k are the FOURIER *coefficients of the functions $f(x)$ and $g(x)$ respectively.*

Proof. The FOURIER coefficients of the sum $f(x) + g(x)$ are the numbers $a_k + b_k$. The PARSEVAL equality for the sum $f(x) + g(x)$ therefore has the appearance:

$$\int_a^b p(f^2 + 2fg + g^2)\,dx = \sum_{k=1}^{\infty} A_k(a_k^2 + 2a_k b_k + b_k^2). \tag{37}$$

Now

$$\int_a^b pf^2\,dx = \sum_{k=1}^{\infty} A_k a_k^2, \quad \int_a^b pg^2\,dx = \sum_{k=1}^{\infty} A_k b_k^2.$$

This and (37) imply (36). The formula (36) is a generalization of the PARSEVAL equality; for $g(x) = f(x)$ it gives the usual PARSEVAL equality.

Theorem 2. (V. A. STEKLOV [1]). *Let A be a set of functions everywhere dense in $L^2_{p(x)}$. If the* PARSEVAL *equality holds for all functions in A, then the underlying orthogonal system is closed.*

Proof. Let $f(x)$ be an arbitrary function of $L^2_{p(x)}$ and suppose $\varepsilon > 0$ is given. Then there exists a function $g(x) \in A$ for which

$$\|f - g\| < \frac{\varepsilon}{2}.$$

Now for $g(x)$ the PARSEVAL equality holds, so that one can find a linear combination $U(x)$ of functions $\omega_k(x)$ such that $\|U - g\| < \frac{\varepsilon}{2}$, and hence

$$\|U - f\| < \varepsilon.$$

The function $f(x)$ can therefore be approximated with arbitrary accuracy by a linear combination of functions of the system; this means however that the PARSEVAL equality holds for $f(x)$.

Corollary 1. *If the* PARSEVAL *equality is satisfied for every arbitrary polynomial, then the system* $\{\omega_k(x)\}$ *is closed.*

Indeed, the class P of all polynomials is everywhere dense in $L^2_{p(x)}$.

Corollary 2. *The trigonometric system* (10) *is closed if the interval of definition has length* 2π.

For the PARSEVAL equality holds for each trigonometric polynomial since it is a linear combination of the functions (10), and the set of these polynomials is dense in $L^2_{p(x)}$.

Definition 2. A system of functions $\Phi = \{\varphi(x)\}$ $(a \leqq x \leqq b)$ is called *complete* if there exists no function except the zero function in $L^2_{p(x)}$ which is orthogonal to all the functions $\varphi(x)$ of Φ.

Theorem 3. *An orthogonal system is complete if and only if it is closed.*[1]

Proof. Suppose the orthogonal system $\{\omega_k(x)\}$ with the weight $p(x)$ is closed. If a function $f(x)$ is orthogonal to all the functions $\omega_k(x)$ of the system, then all the FOURIER coefficients of the function $f(x)$ are equal to zero. The PARSEVAL equality for $f(x)$ therefore takes the form

$$\int_a^b p(x) f^2(x)\, dx = 0,$$

which can be satisfied only by $f(x) = 0$. Therefore, closedness of the orthogonal system implies completeness.

Now suppose that the system $\{\omega_k(x)\}$ is not closed. Then there exists a function $g(x)$ in $L^2_{p(x)}$ for which the PARSEVAL equality does not hold.

If c_k are the FOURIER coefficients of this function, then according to the BESSEL inequality

[1] This simple reciprocal relation between closedness and completeness occurs only in the space $L^2_{p(x)}$; cf. G. M. FICHTENHOLZ [1].

$$\sum_{k=1}^{\infty} A_k c_k^2 < \int_a^b p(x) g^2(x)\, dx.$$

However, from the RIESZ-FISCHER theorem there exists a function $f(x)$ in $L^2_{p(x)}$ for which the c_k are also the FOURIER coefficients as well as for $g(x)$; but as the FOURIER coefficients for $f(x)$

$$\sum_{k=1}^{\infty} A_k c_k^2 = \int_a^b p(x) f^2(x)\, dx.$$

If we put $f(x) - g(x) = r(x)$, then this function, which is not identically zero, is orthogonal to all the functions $\omega_k(x)$, and this implies that the system $\{\omega_k(x)\}$ is not complete.

The theorem is herewith completely proved.

Corollary. *The trigonometric system is complete on an interval of length* 2π.

CHAPTER III

LINEARLY INDEPENDENT SYSTEMS OF FUNCTIONS

§1. Linear Independence. The Gram Determinant. The Schmidt Theorem

Definition 1. A system of functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ defined on the interval $[a, b]$ is called *linearly dependent* if there exist n constants $\alpha_1, \alpha_2, \ldots, \alpha_n$, of which at least one is different from zero, such that

$$\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x) \equiv 0 \tag{38}$$

(as usual, we consider every function equivalent to the zero function to be identically zero). If there does not exist such a set of constants and if (38) implies

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0,$$

then the system of functions is called *linearly independent.*

If a function equivalent to the zero function occurs in the system, then the system is linearly dependent; if any part of the system is linearly dependent, then the entire system is also.

Examples. 1°. *Every finite orthogonal system with weight function $p(x)$ is linearly independent.* For suppose $\{\omega_k(x)\}$ is such a system and

$$\alpha_1\omega_1(x) + \cdots + \alpha_n\omega_n(x) = 0,$$

then multiplying this equation with $p(x)\omega_i(x)$ and integrating we obtain $\alpha_i = 0$.

2°. *If $n_1, n_2, \ldots, n_m$ are whole numbers which are pairwise unequal, then the functions $x^{n_1}, x^{n_2}, \ldots, x^{n_m}$ form a linearly independent system on any interval.* This follows from the fact that a polynomial can have only a finite number of roots.

Definition 2. A denumerably infinite system of functions is called *linearly independent* if every finite subsystem is linearly independent.

For example, the system $1, x, x^2, x^3, \ldots$ is linearly independent.

We now introduce the following notation: if $f(x)$ and $g(x)$ are two functions of $L^2_{p(x)}$, then (f, g) shall denote the integral

$$(f, g) = \int_a^b p(x)f(x)g(x)\,dx.$$

Definition 3. Let the functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ defined on $[a, b]$ belong to $L^2_{p(x)}$. Then the determinant

$$\Delta_n = \begin{vmatrix} (\varphi_1, \varphi_1) & (\varphi_1, \varphi_2) \cdots (\varphi_1, \varphi_n) \\ (\varphi_2, \varphi_1) & (\varphi_2, \varphi_2) \cdots (\varphi_2, \varphi_n) \\ \cdots\cdots\cdots\cdots\cdots \\ (\varphi_n, \varphi_1) & (\varphi_n, \varphi_2) \cdots (\varphi_n, \varphi_n) \end{vmatrix}$$

is called the GRAM[1] *determinant* of the system of functions $\{\varphi_k(x)\}$.

Theorem 1. *A system of functions is linearly dependent if and only if its* GRAM *determinant vanishes.*

Proof. If the system $\{\varphi_k(x)\}$ $(k = 1, 2, \ldots, n)$ is linearly dependent, then there exists a set of constants $\{\alpha_k\}$ not all zero which satisfies (38). Multiplying (38) by $p(x)\varphi_1(x)$, $p(x)\varphi_2(x)$, $\ldots$, $p(x)\varphi_n(x)$ in succession and integrating each time, we obtain the n equations

$$\left.\begin{aligned} \alpha_1(\varphi_1, \varphi_1) + \alpha_2(\varphi_1, \varphi_2) + \cdots + \alpha_n(\varphi_1, \varphi_n) &= 0, \\ \alpha_1(\varphi_2, \varphi_1) + \alpha_2(\varphi_2, \varphi_2) + \cdots + \alpha_n(\varphi_2, \varphi_n) &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \alpha_1(\varphi_n, \varphi_1) + \alpha_2(\varphi_n, \varphi_2) + \cdots + \alpha_n(\varphi_n, \varphi_n) &= 0. \end{aligned}\right\} \tag{39}$$

The numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ therefore form the solution set of a linear homogeneous system of equations with determinant Δ_n. Since not all the α_k are zero, it must follow that

$$\Delta_n = 0.$$

Conversely, suppose $\Delta_n = 0$. Then there exists a set of numbers $\{\alpha_k\}$ not all zero which satisfies the equations (39). We now write (39) in the form

$$\int_a^b p(x)\varphi_1(x)[\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x)]\,dx = 0,$$

$$\int_a^b p(x)\varphi_2(x)[\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x)]\,dx = 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\int_a^b p(x)\varphi_n(x)[\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x)]\,dx = 0.$$

Multiplying the first of these equations by α_1, the second by α_2, etc. and adding the results, we find

[1] GRAM [1].

$$\int_a^b p(x)[\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x)]^2 dx = 0,$$

whence (38) follows, i.e. the system is linearly dependent.

Corollary. *If* $\Delta_n \neq 0$, *then no one of the determinants* $\Delta_1, \Delta_2, \ldots, \Delta_{n-1}$ *is equal to zero.*[1]

For $\Delta_n \neq 0$ implies the linear independence of the system $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$; hence every smaller system $\varphi_1(x), \varphi_2(x), \ldots, \varphi_m(x)$ (with $m < n$) is linearly independent, and therefore from Theorem 1 $\Delta_m \neq 0$.

Lemma. *Let* $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ *be n functions defined on* $[a, b]$ *which belong to* $L^2_{p(x)}$. *Define*

$$\psi_n(x) = \begin{vmatrix} (\varphi_1, \varphi_1) & (\varphi_1, \varphi_2) \cdots (\varphi_1, \varphi_{n-1}) & \varphi_1(x) \\ (\varphi_2, \varphi_1) & (\varphi_2, \varphi_2) \cdots (\varphi_2, \varphi_{n-1}) & \varphi_2(x) \\ \cdots & \cdots & \cdots \\ (\varphi_n, \varphi_1) & (\varphi_n, \varphi_2) \cdots (\varphi_n, \varphi_{n-1}) & \varphi_n(x) \end{vmatrix},$$

then

$$(\psi_n, \varphi_k) = \begin{cases} 0, & \text{if } k < n; \\ \Delta_n, & \text{if } k = n. \end{cases} \tag{40}$$

Proof. Multiply $\psi_n(x)$ by $p(x)\varphi_k(x)$ where in so doing the last column of the determinant is multiplied by this expression. After the multiplication carry out the integration of the product by integrating the last column. The rest is obvious.

Expanding the determinant $\psi_n(x)$ in terms of the elements of the last column, we obtain

$$\psi_n(x) = \alpha_1\varphi_1(x) + \cdots + \alpha_{n-1}\varphi_{n-1}(x) + \Delta_{n-1}\varphi_n(x). \tag{41}$$

If we assume linear independence of the functions $\varphi_k(x)$, then $\Delta_{n-1} \neq 0$, and hence $\psi_n(x)$ is nonzero.

Multiplying (41) by $p(x)\psi_n(x)$, integrating, and taking (40) into account, we obtain

$$\int_a^b p(x)\psi_n^2(x)\, dx = \Delta_{n-1}\Delta_n. \tag{42}$$

The determinant Δ_n therefore has the same sign as Δ_{n-1}. For the same reason, the signs of Δ_{n-1} and Δ_{n-2} agree; continuing in this manner we find that the sign of Δ_n is the same as the sign of $\Delta_1 = (\varphi_1, \varphi_1) > 0$. We therefore have

[1] Δ_1 shall mean (φ_1, φ_1).

Theorem 2. *The* GRAM *determinant of a linearly independent system is strictly positive.*

These considerations can now be used to prove the important "orthogonalization theorem":

Theorem 3. (E. SCHMIDT [1]). *Let $\varphi_1(x), \varphi_2(x), \ldots$ be a finite or denumerably infinite system of functions defined on $[a, b]$ which belong to $L^2_{p(x)}$. Then it is possible to construct an orthonormal system $\omega_1(x), \omega_2(x), \ldots$ which satisfies the following two conditions:*

1. *Each function $\omega_n(x)$ is a linear combination of the first n functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$;*

2. *Each function $\varphi_n(x)$ is a linear combination of the first n functions $\omega_1(x), \omega_2(x), \ldots, \omega_n(x)$.*

Proof. We define

$$\omega_1(x) = \frac{\varphi_1(x)}{\sqrt{\Delta_1}}, \quad \omega_n(x) = \frac{\psi_n(x)}{\sqrt{\Delta_{n-1}\Delta_n}} \qquad (n \geqq 2),$$

where $\psi_n(x)$ is the determinant defined in the lemma.

Since $\psi_n(x)$ is a linear combination of the functions $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$, so is $\omega_n(x)$.

The lemma then ensures the orthogonality of $\psi_n(x)$ to all the functions $\varphi_1, \varphi_2, \ldots, \varphi_{n-1}$ and hence also to any linear combination of them. In particular, $\psi_n(x)$, and therefore also $\omega_n(x)$, is orthogonal to all the functions $\omega_1(x), \omega_2(x), \ldots, \omega_{n-1}(x)$. This means that the system $\{\omega_k(x)\}$ is an orthogonal system (with weight function $p(x)$ to be explicit). From (42) it follows that for $n \geqq 2$

$$\int_a^b p(x)\omega_n^2(x)\,dx = 1,$$

and for $n = 1$ the equation is trivial. The system $\{\omega_k(x)\}$ is therefore orthonormal.

It remains to prove that $\varphi_n(x)$ can also be expressed linearly in terms of $\omega_1(x), \omega_2(x), \ldots, \omega_n(x)$. For $n = 1$ this is obvious. If it has been proved for all $n < m$, then according to (41) we have

$$\varphi_m(x) = \frac{1}{\Delta_{m-1}}\psi_m(x) - \sum_{i=1}^{m-1}\frac{\alpha_i}{\Delta_{m-1}}\varphi_i(x).$$

If we here replace $\psi_m(x)$ by $\sqrt{\Delta_{m-1}\Delta_m}\,\omega_m(x)$ and each function $\varphi_i(x)$ $(i = 1, 2, \ldots, m-1)$ by a linear combination of the functions $\omega_1(x), \ldots, \omega_i(x)$, then $\varphi_m(x)$ is also seen to be a linear combination of the functions $\omega_1(x), \ldots, \omega_m(x)$.

The proof of the theorem has now been completed.

The process of constructing the system $\{\omega_k(x)\}$ is called orthogonalization of the original system $\{\varphi_k(x)\}$.

In view of the fact that a function orthogonal to all the functions $\varphi_k(x)$ is also orthogonal to all the functions $\omega_k(x)$ and vice versa, we obtain

Theorem 4. *The systems* $\{\varphi_k(x)\}$ *and* $\{\omega_k(x)\}$ *are either both complete or else neither is complete.*

§2. Approximation by Linearly Independent Functions

Suppose that a system of linearly independent functions $\varphi_1(x)$, $\varphi_2(x)$, . . . , $\varphi_n(x)$ defined on $[a, b]$ belongs to $L^2_{p(x)}$. We now select any function $f(x) \in L^2_{p(x)}$ and ask the question: What linear combination of the functions $\varphi_k(x)$ provides the best approximation in mean to the function $f(x)$? We have already investigated this question under the assumption that the functions $\varphi_k(x)$ form an orthogonal system with weight function $p(x)$. We now drop this assumption.

After orthogonalization of the system $\{\varphi_k(x)\}$ according to the SCHMIDT procedure, we obtain the system $\{\omega_k(x)\}$. Every linear combination of the functions $\varphi_k(x)$ is then also a linear combination of the functions $\omega_k(x)$, and conversely. However, according to the TOEPLER theorem, among all linear combinations of the functions $\omega_k(x)$ there exists exactly one which minimizes the integral

$$\int_a^b p(x)[f(x) - U(x)]^2\,dx, \tag{43}$$

namely the nth partial FOURIER sum of the function $f(x)$; hence there also exists exactly one linear combination

$$U(x) = \sum_{i=1}^{n} \bar{d}_i\varphi_i(x) \tag{44}$$

of the functions $\varphi_i(x)$ for which the integral (43) becomes a minimum. It is to be noted that not only the combination $U(x)$, but also the coefficients d_i are uniquely determined. For if in addition to the representation (44) $U(x)$ were also represented by

$$U(x) = \sum_{i=1}^{n} \bar{d}_i\varphi_i(x),$$

then

$$\sum_{i=1}^{n} (\bar{d}_i - d_i)\varphi_i(x) = 0,$$

whence because of the linear independence of the functions $\varphi_i(x)$ it immediately follows that $\bar{d}_i = d_i$ $(i = 1, 2, \ldots, n)$.

We now undertake the actual determination of the coefficients d_i and of the minimum value of the integral (43).

Lemma. *The linear combination $U(x)$ which minimizes the integral*

$$\int_a^b p(x)[f(x) - U(x)]^2\,dx$$

has the property that the difference

$$f(x) - U(x) \tag{45}$$

is orthogonal to all the functions $\varphi_k(x)$.

Proof. Let $\{\omega_k(x)\}$ be the orthogonal system that is obtained by the orthogonalization of the system $\{\varphi_k(x)\}$. $U(x)$ is then the nth partial FOURIER sum of the function $f(x)$:

$$U(x) = \sum_{k=1}^{n} c_k\omega_k(x) \qquad \left(c_k = \int_a^b p(x)f(x)\omega_k(x)\,dx\right).$$

It is now sufficient to prove the orthogonality of the difference (45) to the functions $\omega_k(x)$ rather than to the functions $\varphi_k(x)$. This is however obvious, since

$$\int_a^b p(x)\left[f(x) - \sum_{k=1}^{n} c_k\omega_k(x)\right]\omega_i(x)\,dx = c_i - c_i = 0.$$

The lemma is herewith proved.

The coefficients d_i which are sought therefore satisfy the relation

$$\int_a^b p(x)\left[f(x) - \sum_{k=1}^{n} d_k\varphi_k(x)\right]\varphi_i(x)\,dx = 0 \quad (i = 1, 2, \ldots, n),$$

so that they are determined by the system of equations

$$\begin{aligned}
d_1(\varphi_1, \varphi_1) + d_2(\varphi_1, \varphi_2) + \cdots + d_n(\varphi_1, \varphi_n) &= (f, \varphi_1),\\
d_1(\varphi_2, \varphi_1) + d_2(\varphi_2, \varphi_2) + \cdots + d_n(\varphi_2, \varphi_n) &= (f, \varphi_2),\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\\
d_1(\varphi_n, \varphi_1) + d_2(\varphi_n, \varphi_2) + \cdots + d_n(\varphi_n, \varphi_n) &= (f, \varphi_n).
\end{aligned}$$

The determinant of this system is the nonzero GRAM determinant of the linearly independent system $\{\varphi_k(x)\}$, so that there exists only one solution:

$$d_i = \frac{\Delta_n^{(i)}}{\Delta_n},$$

where $\Delta_n^{(i)}$ denotes the determinant obtained from Δ_n by replacing the ith column by the column of constant terms.

The numbers d_i are hereby determined. We now determine the minimum value of the integral (43). This minimum value is

$$\rho_n = \int_a^b p(x)\left[f(x) - \sum_{i=1}^{n} \frac{\Delta_n^{(i)}}{\Delta_n}\varphi_i(x)\right]^2 dx.$$

From this it follows that

$$\rho_n = \int_a^b p(x)\left[f(x) - \sum_{i=1}^{n} \frac{\Delta_n^{(i)}}{\Delta_n}\varphi_i(x)\right] f(x)\,dx$$

$$- \sum_{k=1}^{n} \frac{\Delta_n^{(k)}}{\Delta_n} \int_a^b p(x)\left[f(x) - \sum_{i=1}^{n} \frac{\Delta_n^{(i)}}{\Delta_n}\varphi_i(x)\right]\varphi_k(x)\,dx.$$

According to the lemma the second sum vanishes, so that

$$\rho_n = (f,f) - \sum_{i=1}^{n} \frac{\Delta_n^{(i)}}{\Delta_n}(f,\varphi_i). \tag{46}$$

Expressed in another manner this is

$$\rho_n = \frac{1}{\Delta_n}\left\{(f,f)\,\Delta_n - \sum_{i=1}^{n} \begin{vmatrix} (\varphi_1,\varphi_1)\cdots(\varphi_1,\varphi_{i-1})(\varphi_1,f)(\varphi_1,\varphi_{i+1})\cdots(\varphi_1,\varphi_n) \\ (\varphi_2,\varphi_1)\cdots(\varphi_2,\varphi_{i-1})(\varphi_2,f)(\varphi_2,\varphi_{i+1})\cdots(\varphi_2,\varphi_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (\varphi_n,\varphi_1)\cdots(\varphi_n,\varphi_{i-1})(\varphi_n,f)(\varphi_n,\varphi_{i+1})\cdots(\varphi_n,\varphi_n) \end{vmatrix} (f,\varphi_i)\right\}.$$

As we shall see presently, we obtain for the expression in braces

$$D = \begin{vmatrix} (f,f) & (f,\varphi_1) & (f,\varphi_2) & \cdots (f,\varphi_n) \\ (\varphi_1,f) & (\varphi_1,\varphi_1) & (\varphi_1,\varphi_2)\cdots(\varphi_1,\varphi_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (\varphi_n,f) & (\varphi_n,\varphi_1) & (\varphi_n,\varphi_2)\cdots(\varphi_n,\varphi_n) \end{vmatrix},$$

i.e. the GRAM determinant of the system of functions $f, \varphi_1, \varphi_2, \ldots, \varphi_n$.

Indeed, expanding in terms of elements of the first row we find

$$D = (f,f)\Delta_n + \sum_{i=1}^{n} (-1)^i (f,\varphi_i) \begin{vmatrix} (\varphi_1,f)(\varphi_1,\varphi_1)\cdots(\varphi_1,\varphi_{i-1})\ (\varphi_1,\varphi_{i+1})\cdots(\varphi_1,\varphi_n) \\ (\varphi_2,f)(\varphi_2,\varphi_1)\cdots(\varphi_2,\varphi_{i-1})\ (\varphi_2,\varphi_{i+1})\cdots(\varphi_2,\varphi_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (\varphi_n,f)(\varphi_n,\varphi_1)\cdots(\varphi_n,\varphi_{i-1})(\varphi_n,\varphi_{i+1})\cdots(\varphi_n,\varphi_n) \end{vmatrix}.$$

Moving the first column to the ith place in the subdeterminants thus ob-

tained multiplies the expression under the summation sign by $(-1)^{i-1}$, whereby we obtain precisely the expression in braces above; therefore

$$\rho_n = \frac{D}{\Delta_n}.$$

If we now agree to write the GRAM determinant of a system of functions $f_1, f_2, \ldots, f_n$ in the form

$$\Delta(f_1, f_2, \ldots, f_n)$$

Then our result takes the form

$$\rho_n = \frac{\Delta(\varphi_1, \varphi_2, \ldots, \varphi_n, f)}{\Delta(\varphi_1, \varphi_2, \ldots, \varphi_n)}$$

This expression gives the smallest mean deviation of the linear combination of the system $\{\varphi_k(x)\}$ from the function $f(x)$.

Definition. A system of functions $\Phi = \{\varphi(x)\}$ in $L^2_{p(x)}$ is called a *fundamental system* if the set of all linear combinations of the system is everywhere dense in $L^2_{p(x)}$.

If we are concerned with an orthogonal system, then the concept of the fundamental system clearly coincides with that of the closed system which in turn coincides with the concept of the complete system. However, the same is true in general:

Theorem. *A denumerable, linearly independent system of functions is a fundamental system if and only if it is complete.*

One may even show that denumerability and linear independence of the system are not essential; however, we do not need the theorem in this broader context.

Proof. We first note that the set of linear combinations of the system of functions coincides with the set of linear combinations of those orthonormal functions which we obtain from the given functions by the SCHMIDT orthogonalization procedure. It is therefore sufficient to show that this orthonormal system is fundamental or—what amounts to the same thing—that it is closed. This is the case however if and only if the orthonormal system—and therefore also the original system—is complete.

A system $\{\varphi_k(x)\}$ is certainly fundamental if for each function $f(x) \in L^2_{p(x)}$ the equation

$$\lim_{n\to\infty} \frac{\Delta(f, \varphi_1, \varphi_2, \ldots, \varphi_n)}{\Delta(\varphi_1, \varphi_2, \ldots, \varphi_n)} = 0 \tag{47}$$

holds. It is however sufficient that this equation holds for all functions $f(x)$ of a set A which is everywhere dense in $L^2_{p(x)}$ (since in this case any other

function of $L^2_{p(x)}$ can be approximated to arbitrary accuracy by an element of A, and this latter can further be approximated by a linear combination of the functions $\varphi_k(x)$). Moreover, if (47) is satisfied for all $f(x)$ which belong to a fundamental system, then the system $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$, ... is itself a fundamental system. For example, the system $1, x, x^2, x^3, \ldots$ is a fundamental system since the set of all polynomials is everywhere dense in $L^2_{p(x)}$. If therefore a system $\{\varphi_k(x)\}$ is such that

$$\lim_{n\to\infty} \frac{\Delta(x^m, \varphi_1, \varphi_2, \ldots, \varphi_n)}{\Delta(\varphi_1, \varphi_2, \ldots, \varphi_n)} = 0$$

for $m = 0, 1, 2, 3, \ldots$, then it is a fundamental system.

§3. The Müntz Theorems

Ch. Müntz [1] raised the question: For what whole non-negative exponents $n_1 < n_2 < n_3 < \cdots$ does the sequence of powers $x^{n_1}, x^{n_2}, x^{n_3}, \ldots$ form a fundamental system on the interval [0, 1] (the weight function is here identically one)? As we have just seen, this can be the case if and only if

$$\lim_{s\to\infty} \frac{\Delta(x^m, x^{n_1}, x^{n_2}, \ldots, x^{n_s})}{\Delta(x^{n_1}, x^{n_2}, \ldots, x^{n_s})} = 0 \tag{48}$$

for $m = 0, 1, 2, \ldots$.

If one of the exponents n_i coincides with m, then for $s \geqq i$

$$\Delta(x^m, x^{n_1}, x^{n_2}, \ldots, x^{n_s}) = 0,$$

such that (48) holds. It therefore suffices to consider those values of m which do not coincide with any of the n_i (if there did not exist such values m, this would mean that $\{n_i\}$ contained the entire sequence of natural numbers, and hence $\{x_i^n\}$ would clearly be a fundamental system; we exclude this trivial case from the beginning).

To study the Müntz problem we need the following theorem.

Theorem (Cauchy).

$$\begin{vmatrix} \frac{1}{a_1+b_1} & \frac{1}{a_1+b_2} & \cdots & \frac{1}{a_1+b_n} \\ \frac{1}{a_2+b_1} & \frac{1}{a_2+b_2} & \cdots & \frac{1}{a_2+b_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{a_n+b_1} & \frac{1}{a_n+b_2} & \cdots & \frac{1}{a_n+b_n} \end{vmatrix} = \frac{\prod_{i>k} (a_i - a_k)(b_i - b_k)}{\prod_{i,k} (a_i + b_k)} \tag{49}$$

is an identity.

Proof. Subtract the last row of the determinant from each of the preceding rows; one can then factor the term

$$\frac{\prod_{i=1}^{n-1}(a_n - a_i)}{(a_n + b_1)(a_n + b_2)\cdots(a_n + b_n)}$$

from the determinant. Next subtract the last column from the other columns; this makes it possible to factor the term

$$\frac{\prod_{i=1}^{n-1}(b_n - b_i)}{(a_1 + b_n)(a_2 + b_n)\cdots(a_{n-1} + b_n)}$$

from the determinant, and its order is hereby reduced by one. Repetition of this procedure leads to the identity (49); details of the calculation are left to the reader.

Proceeding now to the MÜNTZ problem, we first note that

$$(x^p, x^q) = \int_0^1 x^{p+q}\,dx = \frac{1}{p+q+1}.$$

From this it follows that

$$\Delta(x^{n_1}, x^{n_2}, \ldots, x^{n_s}) = \begin{vmatrix} \frac{1}{n_1+n_1+1} & \frac{1}{n_1+n_2+1} & \cdots & \frac{1}{n_1+n_s+1} \\ \frac{1}{n_2+n_1+1} & \frac{1}{n_2+n_2+1} & \cdots & \frac{1}{n_2+n_s+1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n_s+n_1+1} & \frac{1}{n_s+n_2+1} & \cdots & \frac{1}{n_s+n_s+1} \end{vmatrix},$$

and therefore according to identity (49)

$$\Delta(x^{n_1}, x^{n_2}, \ldots, x^{n_s}) = \frac{\prod_{i>k}(n_i - n_k)^2}{\prod_{i,k}(n_i + n_k + 1)},$$

and analogously

$$\Delta(x^m, x^{n_1}, x^{n_2}, \ldots, x^{n_s}) = \frac{\prod_{i>k}(n_i - n_k)^2}{\prod_{i,k}(n_i + n_k + 1)} \cdot \frac{\prod_{i=1}^{s}(m - n_i)^2}{\prod_{i=1}^{s}(m + n_i + 1)^2} \cdot \frac{1}{2m+1}.$$

From this it follows that

$$\frac{\Delta(x^m, x^{n_1}, \ldots, x^{n_s})}{\Delta(x^{n_1}, \ldots, x^{n_s})} = \frac{1}{2m+1} \prod_{i=1}^{s}\left(\frac{m - n_i}{m + n_i + 1}\right)^2,$$

so that condition (48) assumes the form

$$\lim_{s\to\infty} \prod_{i=1}^{s} \frac{n_i - m}{n_i + m + 1} = 0. \tag{50}$$

Since we have excluded the case in which m coincides with some of the n_i from the beginning, for every value of s the product following the limit sign in (50) is different from zero. On the other hand, the numbers n_i increase without bound; without loss of generality we may therefore omit factors in which $n_i < m$, since there are only a finite number of such factors, such that in (50) it can be assumed that $n_i > m$.

We can now write (50) in the form

$$\lim_{s\to\infty} \left[\sum_{i=1}^{s} \ln\left(1 - \frac{m}{n_i}\right) - \sum_{i=1}^{s} \ln\left(1 + \frac{m+1}{n_i}\right) \right] = -\infty. \tag{51}$$

If the series

$$\sum_{i=1}^{\infty} \frac{1}{n_i} \tag{52}$$

diverges,[1] then

$$\sum_{i=1}^{\infty} \ln\left(1 - \frac{m}{n_i}\right) = -\infty, \quad \sum_{i=1}^{\infty} \ln\left(1 + \frac{m+1}{n_i}\right) = +\infty, \tag{53}$$

and therefore (51) is satisfied. If however the series (52) converges, then the series (53) also converge, so that (51) is not satisfied. We therefore have

Theorem (CH. MÜNTZ). *A power sequence*

$$x^{n_1}, x^{n_2}, x^{n_3}, \ldots$$

with non-negative integer exponents

$$n_1 < n_2 < n_3 < \cdots$$

is a fundamental system in $L^2([0, 1])$ *if and only if the series*

$$\sum_{i=1}^{\infty} \frac{1}{n_i}$$

diverges.

Closely related to this result is a second theorem of the same author which contains the solution to the following problem: What property must a non-negative integer sequence of exponents $\{n_i\}$ $(n_0 < n_1 < n_2 < \cdots)$ have in order that any function of $C([0, 1])$ can be uniformly approximated by polynomials of the form

[1] In case $n_1 = 0$ we of course begin the summation in series (52) with $i = 2$.

$$\sum_{i=0}^{s} c_i x^{n_i} \tag{54}$$

If a sequence $\{n_i\}$ has this property, the power sequence $\{x^{n_i}\}$ is called a fundamental system in $C([0, 1])$.

Theorem (CH. MÜNTZ). *A power sequence* $\{x^{n_i}\}$ $(n_0 < n_1 < n_2 < \cdots)$ *is a fundamental system in* $C([0, 1])$ *if and only if* 1) $n_0 = 0$ *and* 2) *the series*

$$\sum_{i=1}^{\infty} \frac{1}{n_i} \tag{55}$$

diverges.

Proof. Necessity of the condition $n_0 = 0$ follows by simply noting that if $n_0 > 0$ the polynomial (54) is itself zero at $x = 0$ and cannot therefore approximate those continuous functions for which $f(0) \neq 0$. Further, if a power sequence is fundamental in $C([0, 1])$, then it is fundamental a fortiori [1] in $L^2([0, 1])$, so that according to the previous theorem the series (55) diverges.

To prove sufficiency we assume that both conditions are satisfied.

If $i \geqq 2$, then $n_i > 1$. Divergence of the series (55) implies divergence of

$$\sum_{i=2}^{\infty} \frac{1}{n_i - 1};$$

the power sequence $\{x^{n_i - 1}\}$ is therefore a fundamental system in $L^2([0, 1])$ for $i \geqq 2$ and thus also for $i \geqq 1$.

After this preliminary remark we choose an arbitrary natural number m and a real number $\varepsilon > 0$. Since the function x^{m-1} belongs to $L^2([0, 1])$, it is possible to find coefficients a_i which satisfy the inequality

$$\int_0^1 \left[x^{m-1} - \sum_{i=1}^{s} a_i x^{n_i - 1} \right]^2 dx < \frac{\varepsilon^2}{m^2}.$$

Now if $0 \leqq x \leqq 1$, the SCHWARZ inequality implies

$$\int_0^x \left| x^{m-1} - \sum_{i=1}^{s} a_i x^{n_i - 1} \right| dx < \frac{\varepsilon}{m};$$

and from this it follows that

$$\left| x^m - \sum_{i=1}^{s} \frac{m}{n_i} a_i x^{n_i} \right| < \varepsilon.$$

Therefore for $m > 0$ the power x^m can be approximated to arbitrary accuracy by means of the polynomials (54). Since however $n_0 = 0$, the condition

[1] Since $C([0, 1])$ is everywhere dense in L^2 $([0, 1])$.

$m > 0$ is unnecessary, and therefore any polynomial can be approximated to arbitrary accuracy by means of the polynomials (54). The rest of the proof consists of invoking the WEIERSTRASS theorem according to which any function of $C([0, 1])$ can be approximated to arbitrary accuracy by means of polynomials.

CHAPTER IV

GENERAL PROPERTIES OF ORTHOGONAL POLYNOMIALS

§ 1. Foundations

The power sequence $\{x^k\}$ $(k = 0, 1, 2, \ldots)$ is linearly independent in every interval $[a, b]$; according to the theorem of E. SCHMIDT it can therefore be orthogonalized for every weight function. This we shall now study more closely.

The GRAM determinant of the original system of functions plays an essential role in the SCHMIDT theorem. If we now introduce the notation

$$\int_a^b p(x)x^n\,dx = \mu_n \qquad (n = 0, 1, 2, \ldots)$$

(where these numbers are called *moments* of the weight function $p(x)$), then

$$(x^p, x^q) = \int_a^b p(x)x^{p+q}\,dx = \mu_{p+q}.$$

The GRAM determinant [1]

$$\Delta_n = \begin{vmatrix} (1,1) & (1,x) & \cdots & (1,x^n) \\ (x,1) & (x,x) & \cdots & (x,x^n) \\ \cdots & \cdots & \cdots & \cdots \\ (x^n,1) & (x^n,x) & \cdots & (x^n,x^n) \end{vmatrix}$$

then becomes

$$\Delta_n = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix}. \tag{56}$$

The function $\psi_n(x)$ of the lemma in § 1 of Chapter III now has the appearance

$$\psi_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} & 1 \\ \mu_1 & \mu_2 & \cdots & \mu_n & x \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} & x^n \end{vmatrix},$$

[1] This determinant is of order $n + 1$. There is hence a difference (although a minor one) between our present notation and that formerly used where Δ_n was of order n.

so that the functions $\omega_n(x)$ which form the orthonormal system are now

$$\left.\begin{aligned} \omega_0(x) &= \frac{1}{\sqrt{\Delta_0}} \quad (\Delta_0 = \mu_0), \\ \omega_n(x) &= \frac{1}{\sqrt{\Delta_{n-1}\Delta_n}} \begin{vmatrix} \mu_0 & \mu_1 & \cdots \mu_{n-1} & 1 \\ \mu_1 & \mu_2 & \cdots \mu_n & x \\ \cdot & \cdot & \cdots & \cdot \\ \mu_n & \mu_{n+1} & \cdots \mu_{2n-1} & x^n \end{vmatrix} \quad (n = 1, 2, \ldots) \end{aligned}\right\}. \tag{57}$$

We hereby note that $\omega_n(x)$ is a polynomial of precisely nth degree, since its coefficient of x^n has the value

$$\sqrt{\frac{\Delta_{n-1}}{\Delta_n}} \neq 0$$

Therefore in our case the SCHMIDT theorem can be restated as follows:

Theorem 1. *For each weight function $p(x)$ defined on the interval $[a, b]$ there exists an orthonormal system of polynomials on $[a, b]$*

$$\omega_0(x), \omega_1(x), \omega_2(x), \ldots \tag{58}$$

with the weight function $p(x)$, where $\omega_n(x)$ is a polynomial of precisely nth degree.

As we have seen, the system (58) can be defined by formula (57). The question is immediately raised if there exists another orthonormal system with weight $p(x)$ which is not formed by means of formula (57). There is clearly no complete uniqueness, since the orthonormality of the system is retained when one or more of the functions of the system (58) is multiplied by -1. If however we fix the sign [2] of the leading coefficient of these polynomials and require in addition that the nth polynomial of the system should be of exactly nth degree, then we obtain a unique determination of the system (58). For the proof we need a simple

Lemma. *Let $Q_0(x), Q_1(x), Q_2(x), \ldots$ be a polynomial system in which each polynomial $Q_n(x)$ is of exactly nth degree. Then every polynomial of degree $m \geqq 0$ can be represented uniquely in the form*

$$P(x) = \alpha_0 Q_0(x) + \alpha_1 Q_1(x) + \cdots + \alpha_m Q_m(x). \tag{59}$$

Proof. If

$$\begin{aligned} Q_n(x) &= q_0^{(n)} + q_1^{(n)} x + \cdots + q_n^{(n)} x^n \qquad (q_n^{(n)} \neq 0), \\ P(x) &= p_0 + p_1 x + \cdots + p_m x^m, \end{aligned}$$

then equation (59) is satisfied if and only if

[2] The leading coefficient in the polynomial (57) is positive.

$$\alpha_m q_m^{(m)} = p_m,$$
$$\alpha_{m-1} q_{m-1}^{(m-1)} + \alpha_m q_{m-1}^{(m)} = p_{m-1},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\alpha_0 q_0^{(0)} + \alpha_1 q_0^{(1)} + \cdots + \alpha_m q_0^{(m)} = p_0.$$

This system of equations gives the coefficients $\alpha_m, \alpha_{m-1}, \ldots, \alpha_0$ successively (and of course uniquely).

Theorem 2. *If the polynomial system*

$$\varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots$$

is orthonormal with weight $p(x)$ *where each polynomial* $\varphi_n(x)$ *is exactly of nth degree and in each polynomial* $\varphi_n(x)$ *the leading coefficient is positive, then*

$$\varphi_n(x) = \omega_n(x),$$

where $\omega_n(x)$ *is defined by formula* (57).

Proof. According to the lemma

$$\varphi_n(x) = \alpha_0\omega_0(x) + \alpha_1\omega_1(x) + \cdots + \alpha_n\omega_n(x). \tag{60}$$

But from the same lemma the polynomials $\omega_0(x), \omega_1(x), \ldots, \omega_{n-1}(x)$ can also be linearly expressed in terms of $\varphi_0(x), \varphi_1(x), \ldots, \varphi_{n-1}(x)$ and are therefore orthogonal to $\varphi_n(x)$. Multiplying (60) by $p(x)\omega_0(x)$, $p(x)\omega_1(x)$, $\ldots$, $p(x)\omega_{n-1}(x)$ in succession and integrating each time, we obtain

$$\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1} = 0.$$

Hence $\varphi_n(x) = \alpha_n\omega_n(x)$. From this it follows that

$$\int_a^b p(x)\varphi_n^2(x)\,dx = \alpha_n^2 \int_a^b p(x)\omega_n^2(x)\,dx.$$

Now however both integrals are equal to one, and therefore $\alpha_n = \pm 1$. The theorem is herewith proved.

The lemma which was just previously proved leads to another important result. Every polynomial can now be represented as a linear combination of the polynomials (58). Therefore the PARSEVAL equality holds for every polynomial, and hence on the basis of the STEKLOV theorem we have

Theorem 3. *The orthonormal system* (58) *is closed.*

Corollary. *The system* $1, x, x^2, x^3, \ldots$ *is complete.*

In view of the considerations at the end of § 2 of Ch. III this corollary is trivial.

Since the set H_n of all polynomials of degree at most n is identical with the

set of all linear combinations of the polynomials $\omega_0(x), \omega_1(x), \ldots, \omega_n(x)$, the generalized TOEPLER theorem of Chapter II leads to

Theorem 4. *If* $f(x) \in L^2_{p(x)}([a, b])$ *and* $P(x) \in H_n$, *then the integral*

$$\int_a^b p(x)[f(x) - P(x)]^2\, dx \tag{61}$$

is a minimum if and only if $P(x)$ *is chosen equal to the partial* FOURIER *sum*

$$S_n(x) = \sum_{k=0}^{n} c_k \omega_k(x) \qquad \left(c_k = \int_a^b p(x) f(x) \omega_k(x)\, dx\right).$$

The problem of the best approximation of a given function by means of polynomials is therefore considerably simplified by adopting the integral (61) as a measure of the deviation of the polynomial from the function rather than taking

$$\max |f(x) - P(x)|,$$

as in the first part of the book. Now let us find the minimum value of the integral (61). From the general formula (28)

$$\int_a^b p(x)[f(x) - S_n(x)]^2\, dx = \int_a^b p(x) f^2(x)\, dx - \sum_{k=0}^{n} c_k^2.$$

On the other hand, since the system (58) is closed

$$\int_a^b p(x) f^2(x)\, dx = \sum_{k=0}^{\infty} c_k^2.$$

hence

$$\int_a^b p(x)[f(x) - S_n(x)]^2\, dx = \sum_{k=n+1}^{\infty} c_k^2.$$

The polynomials $\omega_n(x)$ have a second property in addition to the extremal property expressed in Theorem 4:

Theorem 5. *If* $P(x)$ *is a polynomial of degree* n *whose leading coefficient has the value one, then the integral*

$$\int_a^b p(x) P^2(x)\, dx \tag{62}$$

is a minimum if and only if [3]

[3] Conditions other than fixing the leading coefficient can be imposed on the coefficients of $P(x)$. See, for example, J. SHOHAT [1] and J. L. GERONIMUS [1].

$$P(x) = \sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\,\omega_n(x). \tag{63}$$

Proof. According to the lemma, every polynomial of degree n with leading coefficient equal to one has the form

$$P(x) = \alpha_0\omega_0(x) + \cdots + \alpha_{n-1}\omega_{n-1}(x) + \sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\,\omega_n(x), \tag{64}$$

and conversely every such linear combination represents a polynomial of degree n with leading coefficient one. The choice of the polynomial $P(x)$ is therefore equivalent to the choice of the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$. For the polynomial (64) the PARSEVAL equality now gives

$$\int_a^b p(x)P^2(x)\,dx = \sum_{k=0}^{n-1}\alpha_k^2 + \frac{\Delta_n}{\Delta_{n-1}};$$

We clearly obtain the minimum value of this integral if and only if

$$\alpha_0 = \alpha_1 = \cdots = \alpha_{n-1} = 0.$$

We now drop the normalization and concern ourselves with orthogonal polynomial systems as such. It turns out that such a system is uniquely determined when the leading coefficient of each polynomial of the system is given.

Theorem 6. *Let $\varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots$ be an orthogonal system with weight $p(x)$ in which each element $\varphi_n(x)$ is a polynomial of degree n. If the leading coefficient of $\varphi_n(x)$ is K_n $(n = 0, 1, 2, \ldots)$, then necessarily*

$$\varphi_0(x) = K_0, \quad \varphi_n(x) = K_n\sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\,\omega_n(x) \qquad (n \geqq 1). \tag{65}$$

Proof. Clearly $\varphi_0(x) = K_0$. For $n \geqq 1$ the expression for $\varphi_n(x)$ is obtained as in Theorem 2 on the basis of the orthogonality of $\varphi_n(x)$ to $\omega_0(x)$, $\omega_1(x), \ldots, \omega_{n-1}(x)$. For the special case $K_n = 1$ $(n = 0, 1, 2, \ldots)$ we shall use the notation [4]

$$\tilde{\omega}_0(x) = 1, \quad \tilde{\omega}_n(x) = \sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\,\omega_n(x) \tag{66}$$

In addition, from the formulas (65) one obtains

$$A_0 = \int_a^b p(x)\varphi_0^2(x)\,dx = K_0^2\Delta_0,$$

[4] As far as I know, this practical notation was first introduced by V. L. GONCHAROV.

$$A_n = \int_a^b p(x)\varphi_n^2(x)\,dx = K_n^2 \frac{\Delta_n}{\Delta_{n-1}} \tag{67}$$

(we remind the reader that the numbers Δ_n occur in the expressions for the FOURIER coefficients of an unnormalized orthogonal system; they were defined in equation (12)).

Finally, we note that for the actual construction of the system (58), finding the numbers Δ_n, and etc. it is not necessary to know the weight function $p(x)$ itself, but rather only its moments μ_n; these alone enter formulas (56) and (57)

§2. Roots of Orthogonal Polynomials. A Recursion Formula

Theorem 1. *All the roots of the polynomial $\omega_n(x)$ are real, simple, and contained in the interval (a, b).*

Proof. We first assume that $\omega_n(x)$ has no root of odd order in the interval (a, b). Then the polynomial $\omega_n(x)$ does not change sign on the interval $[a, b]$ and hence the equality

$$\int_a^b p(x)\omega_n(x)\,dx = 0,$$

which expresses the orthogonality of $\omega_n(x)$ to $\omega_0(x) = \text{const.}$ is impossible. Therefore, there exists at least one root of odd order in the interval (a, b). Let the number of roots of odd order be $r < n$, and denote the roots themselves by $\xi_1, \xi_2, \ldots, \xi_r$.

Define

$$Q(x) = (x - \xi_1)(x - \xi_2)\cdots(x - \xi_r).$$

The polynomial $Q(x)$ is of rth degree and is therefore orthogonal to $\omega_n(x)$ (it is a linear combination of $\omega_0, \omega_1, \ldots, \omega_r$). Therefore

$$\int_a^b p(x)Q(x)\omega_n(x)\,dx = 0;$$

This however is impossible, since all the roots of the polynomial $Q(x)\omega_n(x)$ are of even order, and hence this polynomial does not change sign on $[a, b]$. We must therefore have $r = n$, and everything is herewith proved.

Theorem 2. *The three successive polynomials $\tilde{\omega}_{n+2}(x)$, $\tilde{\omega}_{n+1}(x)$, $\tilde{\omega}_n(x)$ are related to one another by a recursion formula*

$$\tilde{\omega}_{n+2}(x) = (x - \alpha_{n+2})\tilde{\omega}_{n+1}(x) - \lambda_{n+1}\tilde{\omega}_n(x) \qquad (n = 0, 1, 2, \ldots) \tag{68}$$

where α_{n+2} and λ_{n+1} are certain constants.

Recall that the polynomials $\tilde{\omega}_n(x)$ defined by the formula (66) have the leading coefficient equal to one.

Proof. Consider the product $x\tilde{\omega}_{n+1}(x)$. Since it is a polynomial of degree $n+2$, it can be represented in the form

$$x\tilde{\omega}_{n+1}(x) = c_0\tilde{\omega}_0(x) + c_1\tilde{\omega}_1(x) + \cdots + c_{n+2}\tilde{\omega}_{n+2}(x). \tag{69}$$

Comparison of the two leading coefficients gives $c_{n+2} = 1$. We then multiply equation (69) by $p(x)\tilde{\omega}_k(x)$ where $k < n$ and integrate the product. Since the degree of the polynomial $x\tilde{\omega}_k(x)$ is less than $n+1$, the left-hand side gives zero. On the right-hand side only the term

$$c_k \int_a^b p(x)\tilde{\omega}_k^2(x)\,dx$$

survives, while all others vanish due to the orthogonality of the system $\{\tilde{\omega}_n(x)\}$; therefore $c_k = 0$. We now have

$$c_0 = c_1 = \cdots = c_{n-1} = 0,$$

and hence equation (69) now assumes the form

$$x\tilde{\omega}_{n+1}(x) = c_n\tilde{\omega}_n(x) + c_{n+1}\tilde{\omega}_{n+1}(x) + \tilde{\omega}_{n+2}(x)$$

which is identical to (68) with the identifications $c_n = \lambda_{n+1}$ and $c_{n+1} = \alpha_{n+2}$.

Substituting in (68) the expressions (66) for the polynomial $\tilde{\omega}_n(x)$, we obtain a recursion formula [5] for the polynomials (58):

$$\sqrt{\frac{\Delta_{n+2}}{\Delta_{n+1}}}\omega_{n+2}(x) = (x - \alpha_{n+2})\sqrt{\frac{\Delta_{n+1}}{\Delta_n}}\omega_{n+1}(x) - \lambda_{n+1}\sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\omega_n(x). \tag{70}$$

Since this formula is somewhat unwieldy, we shall subsequently prefer to use formula (68).

The two numbers α_{n+2} and λ_{n+1} occurring in (68) are easily determined. Multiplying (68) by $p(x)\tilde{\omega}_{n+1}(x)$ and integrating, we obtain

$$\int_a^b p(x)(x - \alpha_{n+2})\tilde{\omega}_{n+1}^2(x)\,dx = 0,$$

whence

$$\alpha_{n+2} = \frac{\int_a^b p(x)x\tilde{\omega}_{n+1}^2(x)\,dx}{\int_a^b p(x)\tilde{\omega}_{n+1}^2(x)\,dx}. \tag{71}$$

[5] One defines $\Delta_{-1} = 1$ in order that formula (70) may also be used for the case $n = 0$.

The numerator of this fraction becomes larger when we replace the factor x in the integrand by its greatest value b. Therefore $\alpha_{n+2} < b$, and analogously $\alpha_{n+2} > a$ or

$$a < \alpha_{n+2} < b \qquad (n = 0, 1, 2, \ldots). \tag{72}$$

To determine λ_{n+1} we multiply (68) by $p(x)\tilde{\omega}_n(x)$ and integrate. The result is

$$\lambda_{n+1}\int_a^b p(x)\tilde{\omega}_n^2(x)\,dx = \int_a^b p(x)x\tilde{\omega}_n(x)\tilde{\omega}_{n+1}(x)\,dx.$$

Now the product $x\tilde{\omega}_n(x)$ can be written in the form

$$x\tilde{\omega}_n(x) = \tilde{\omega}_{n+1}(x) + R(x),$$

where $R(x)$ is a polynomial of degree at most n, and therefore

$$\int_a^b p(x)R(x)\tilde{\omega}_{n+1}(x)\,dx = 0.$$

We hereby obtain

$$\lambda_{n+1} = \frac{\int_a^b p(x)\tilde{\omega}_{n+1}^2(x)\,dx}{\int_a^b p(x)\tilde{\omega}_n^2(x)\,dx}, \tag{73}$$

where clearly

$$\lambda_{n+1} > 0 \tag{74}$$

for all n.

This is of great importance for all that follows. Applying the formulas (67) to (73), we find

$$\lambda_{n+1} = \frac{\Delta_{n-1}\Delta_{n+1}}{\Delta_n^2}, \tag{75}$$

a relation which is also valid for $n = 0$ if we define $\Delta_{-1} = 1$.

The equation

$$\int_a^b p(x)x\tilde{\omega}_n(x)\tilde{\omega}_{n+1}(x)\,dx = \int_a^b p(x)\tilde{\omega}_{n+1}^2(x)\,dx$$

leads us to a useful estimate of λ_{n+1}. Denoting the greater of the two numbers $|a|$ and $|b|$ by C by defining

$$C = \max\{|a|, |b|\},$$

we obtain

$$\int_a^b p(x)\bar{\omega}_{n+1}^2(x)\,dx \leqq C\int_a^b p(x)|\bar{\omega}_n(x)|\,|\bar{\omega}_{n+1}(x)|dx.$$

The SCHWARZ inequality implies

$$\int_a^b p(x)|\bar{\omega}_n(x)|\,|\bar{\omega}_{n+1}(x)|dx \leqq \sqrt{\int_a^b p(x)\bar{\omega}_n^2(x)\,dx}\,\sqrt{\int_a^b p(x)\bar{\omega}_{n+1}^2(x)\,dx},$$

and therefore

$$\int_a^b p(x)\omega_{n+1}^2(x)\,dx \leqq C\sqrt{\int_a^b p(x)\bar{\omega}_n^2(x)\,dx}\,\sqrt{\int_a^b p(x)\bar{\omega}_{n+1}^2(x)\,dx};$$

From this and from (73) it follows that

$$\lambda_{n+1} \leqq C^2.$$

Two consequences can be obtained from Theorem 2:

I. *The two successive polynomials* $\omega_{n+2}(x)$ *and* $\omega_{n+1}(x)$ *have no common roots.*

For such a common root would also be a root of $\omega_n(x)$, and hence a root of $\omega_{n-1}(x)$, and by continuing the procedure it would finally also be a root of $\omega_0(x)$ which is impossible, since $\omega_0(x)$ is a nonzero constant.

II. *If* x_0 *is a root of the polynomial* $\omega_{n+1}(x)$, *then the numbers* $\omega_{n+2}(x_0)$ *and* $\omega_n(x_0)$ *have opposite signs.*

For formula (70) yields

$$\sqrt{\frac{\Delta_{n+2}}{\Delta_{n+1}}}\,\omega_{n+2}(x_0) = -\lambda_{n+1}\sqrt{\frac{\Delta_n}{\Delta_{n-1}}}\,\omega_n(x_0),$$

whence in conjunction with (74) the assertion follows.

Theorem 3. *For* $n > 0$, *the roots of* $\omega_n(x)$ *and* $\omega_{n+1}(x)$ *alternate.*

More precisely: the roots $x_k^{(n)}$ of the polynomial $\omega_n(x)$ and the roots $x_k^{(n+1)}$ of the polynomial $\omega_{n+1}(x)$ satisfy the inequalities

$$a < x_1^{(n+1)} < x_1^{(n)} < x_2^{(n+1)} < \cdots < x_n^{(n+1)} < x_n^{(n)} < x_{n+1}^{(n+1)} < b,$$

such that between any two successive roots of $\omega_{n+1}(x)$ there is exactly one root of $\omega_n(x)$.

Proof. The proof proceeds by induction. First, for $n = 1$ and the single root $x_1^{(1)}$ of the polynomial $\omega_1(x)$

$$a < x_1^{(1)} < b.$$

The polynomial $\omega_0(x)$ is a positive constant. Therefore from consequence II of the preceding theorem $\omega_2(x_1^{(1)})$ is negative. On the other hand $\omega_2(a)$ and $\omega_2(b)$ are positive values, for the roots $x_1^{(2)}$ and $x_2^{(2)}$ of the polynomial $\omega_2(x)$ lie in the interior of the interval (a, b), and hence the signs of $\omega_2(a)$

and $\omega_2(b)$ agree with the signs of this function at $x = -\infty$ and $x = +\infty$ respectively which are both positive, since $\omega_2(x)$ is a polynomial of second degree with a positive leading coefficient.

Each of the two intervals $[a, x_1^{(1)}]$ and $[x_1^{(1)}, b]$ thus has the property that $\omega_2(x)$ has different signs at the two end points. From this it immediately follows that

$$a < x_1^{(2)} < x_1^{(1)} < x_2^{(2)} < b,$$

which proves the theorem for $n = 1$.

Continuing the proof, we first note that the inequality $\omega_n(b) > 0$ holds for all n, while $\omega_n(a)$ is positive for even and negative for odd n. This is based on the same considerations as those employed for $\omega_2(x)$.

We now assume that the theorem has been proved for the index n, and we consider the points

$$a, \quad x_1^{(n+1)}, \quad x_2^{(n+1)}, \quad x_3^{(n+1)}, \quad \ldots, \quad x_{n+1}^{(n+1)}, \quad b.$$

The polynomials $\omega_{n+2}(x)$ and $\omega_n(x)$ have the same signs at the point a and opposite signs at the point $x_1^{(n+1)}$. Now the polynomial $\omega_n(x)$ has no root in the interval $(a, x_1^{(n+1)})$, so that $\omega_n(x_1^{(n+1)})$ has the same sign as $\omega_n(a)$. Therefore, the polynomial $\omega_{n+2}(x)$ changes sign between $x = a$ and $x = x_1^{(n+1)}$ and therefore certainly has a root in the interval $(a, x_1^{(n+1)})$.

At the point $x_2^{(n+1)}$ the polynomials $\omega_{n+1}(x)$ and $\omega_n(x)$ again have different signs. Now the polynomial $\omega_n(x)$ has exactly one (simple) root in the interval $(x_1^{(n+1)}, x_2^{(n+1)})$. It therefore changes sign between $x = x_1^{(n+1)}$ and $x = x_2^{(n+1)}$, whence it follows that $\omega_{n+2}(x)$ also changes sign between these two points and therefore has a root in the interval $(x_1^{(n+1)}, x_2^{(n+1)})$.

By continuing this procedure we convince ourselves that $\omega_{n+2}(x)$ has a root in each interval

$$(a, x_1^{(n+1)}), \quad (x_1^{(n+1)}, x_2^{(n+1)}), \quad \ldots, \quad (x_{n+1}^{(n+1)}, b);$$

however since the number of these intervals is $n + 2$, i.e. the same as the number of roots of $\omega_{n+2}(x)$, each interval contains exactly one root. Therefore the validity of the theorem for n implies that the theorem also holds for $n + 1$. The theorem has now been proved.

The method of the proof can be clarified by the following table which is applicable for any even value of n:

	a	$x_1^{(n+1)}$	$x_2^{(n+1)}$	$x_3^{(n+1)}$	$\cdots$	$x_n^{(n+1)}$	$x_{n+1}^{(n+1)}$	b
Sign $\omega_n(x)$	+	+	−	+	$\cdots$	−	+	+
Sign $\omega_{n+2}(x)$	+	−	+	−	$\cdots$	+	−	+

The polynomials $\omega_n(x)$ have still another important algebraic property which relates them to the STURM functions and serves to determine the number of real roots of a polynomial.

Let

$$\sigma_0, \sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n \tag{76}$$

be any real, nonzero numbers. We now count how often two neighboring numbers σ_i and σ_{i+1} occurring in this sequence have different signs and call the number so obtained the *number of alternations* in the sequence (76). Expressed in another manner: the number of alternations is equal to the number of *negative* products of the form

$$\sigma_i\sigma_{i+1} \qquad (i = 0, 1, 2, \ldots, n-1).$$

If zeros are also admitted in the sequence (76), we delete these and determine the number of alternations in the remaining sequence; we call the number so obtained the number of alternations in the original sequence (76). Thus, for example, in the sequence

$$0, 2, 3, -1, 0, 0, 5, -2$$

the number of alternations is determined from those occurring in the sequence

$$2, 3, -1, 5, -2,$$

and hence is equal to three.

Since in the sequence

$$\omega_0(a), \omega_1(a), \omega_2(a), \ldots, \omega_n(a)$$

two adjacent numbers always have opposite signs, the number of alternations is equal to n. On the other hand, the number of alternations in the sequence

$$\omega_0(b), \omega_1(b), \omega_2(b), \ldots, \omega_n(b)$$

is equal to zero. If in general we denote the number of alternations in the sequence

$$\omega_0(x), \omega_1(x), \omega_2(x), \ldots, \omega_n(x) \tag{77}$$

by $\lambda(x)$, then $\lambda(a) = n$ and $\lambda(b) = 0$.

Theorem 4. *The number of roots of the polynomial $\omega_n(x)$ in the half open interval $(\alpha, \beta]$ is exactly equal to the difference*

$$\lambda(\alpha) - \lambda(\beta).$$

Proof. We first assume that at the points $x = \alpha$ and $x = \beta$ none of the polynomials vanishes. We then enumerate all the points $z_1, z_2, \ldots, z_m$ of the interval (α, β) at which at least one of the polynomials (77) vanishes.

If in each interval (z_i, z_{i+1}) we now choose a point y_i $(i = 1, 2, \ldots, m-1)$ and put $y_0 = \alpha$ and $y_m = \beta$, then

$$\lambda(\alpha) - \lambda(\beta) = \sum_{i=0}^{m-1} [\lambda(y_i) - \lambda(y_{i+1})].$$

We now study the difference $\lambda(y_i) - \lambda(y_{i+1})$ more closely. None of the polynomials (77) assumes the value zero at the points y_i and y_{i+1}, but the interval (y_i, y_{i+1}) contains the point z_{i+1} which is a root of one or more of the polynomials of (77) and which is moreover the only such point in this interval. Let us assume that

$$\omega_n(z_{i+1}) \neq 0,$$

but

$$\omega_{k_1}(z_{i+1}) = \omega_{k_2}(z_{i+1}) = \cdots = \omega_{k_r}(z_{i+1}) = 0.$$

To find $\lambda(y_i)$ we must form all products

$$\omega_0(y_i)\omega_1(y_i),\ \omega_1(y_i)\omega_2(y_i),\ \ldots,\ \omega_{n-1}(y_i)\omega_n(y_i)$$

and count the number of them which are negative; $\lambda(y_{i+1})$ is found analogously.

However, between $x = y_i$ and $x = y_{i+1}$ only the polynomials $\omega_{k_1}(x)$, $\omega_{k_2}(x)$, $\ldots$, $\omega_{k_r}(x)$ change sign. Therefore only products of the form

$$\omega_{k_1-1}(x)\omega_{k_1}(x),\ \omega_{k_1}(x)\omega_{k_1+1}(x),\ \omega_{k_2-1}(x)\omega_{k_2}(x),\ \ldots,\ \omega_{k_r}(x)\omega_{k_r+1}(x)$$

can change their signs. Each of the polynomials $\omega_{k_s}(x)$ occurs in two such products.

Now none of the numbers $k_1 \pm 1$, $k_2 \pm 1$, $\ldots$, $k_r \pm 1$ are among the numbers $k_1, k_2, \ldots, k_r$, since neighboring polynomials have no common roots. If on the other hand j is different from $k_1, k_2, \ldots, k_r$, then the three numbers

$$\omega_j(y_i),\ \omega_j(z_{i+1}),\ \omega_j(y_{i+1})$$

have one and the same sign. Since however

$$\omega_{k_1-1}(z_{i+1})\omega_{k_1+1}(z_{i+1}) < 0,$$

exactly one of the two products

$$\omega_{k_1-1}(y_i)\omega_{k_1}(y_i),\ \omega_{k_1}(y_i)\omega_{k_1+1}(y_i)$$

is negative. In the same way, exactly one of the two products

$$\omega_{k_1-1}(y_{i+1})\omega_{k_1}(y_{i+1}),\ \omega_{k_1}(y_{i+1})\omega_{k_1+1}(y_{i+1})$$

is negative.

Hence it makes no difference if we count the alternations at y_i or at y_{i+1}, since the number of negative products which contain the factor $\omega_{k_1}(x)$ is

equal to one, and the same holds for the other polynomials $\omega_{k_2}(x), \ldots, \omega_{k_r}(x)$. The number of negative products

$$\omega_k(x)\omega_{k+1}(x) \quad (k = 0, 1, 2, \ldots, n-1) \tag{78}$$

therefore does not change between $x = y_i$ and $x = y_{i+1}$, and hence

$$\lambda(y_i) = \lambda(y_{i+1}).$$

We now assume that

$$\omega_n(z_{i+1}) = 0.$$

Other polynomials $\omega_{k_1}(x), \omega_{k_2}(x), \ldots, \omega_{k_r}(x)$ besides the polynomial $\omega_n(x)$ can also vanish at the point z_{i+1}; however, we convince ourselves as above that the number of negative products (78) containing these polynomials as factors does not change between $x = y_i$ and $x = y_{i+1}$.

The product

$$\omega_{n-1}(x)\omega_n(x) \tag{79}$$

is however negative at $x = y_i$ and positive at $x = y_{i+1}$. For the roots of the polynomials $\omega_{n-1}(x)$ and $\omega_n(x)$ alternate with one another, such that to the left of the point z_{i+1} both polynomials have the same number of roots.

The product (79) changes sign at each of these roots; therefore, between $-\infty$ and y_i an *even number* of alternations occurs, and since the product is negative at $x = -\infty$ it is also negative at $x = y_i$. At the point z_{i+1} it again changes sign and becomes positive.

Therefore, in the case under consideration we have the equality

$$\lambda(y_i) - \lambda(y_{i+1}) = 1,$$

which implies that the difference $\lambda(\alpha) - \lambda(\beta)$ is equal to the number of roots z_{i+1} of $\omega_n(x)$.

We must now consider the case in which one or both of the points α, β are roots of one or more of the polynomials (77).

Suppose first that α is such a "bad" point. We then choose an $h > 0$ such that in the interval $(\alpha, \alpha + h]$ there are no roots of the polynomials (77). From what has just been proved the number of roots of $\omega_n(x)$ in the interval $(\alpha + h, \beta)$ (or—what amounts to the same thing here—in the interval $[\alpha + h, \beta]$ is then equal to the difference

$$\lambda(\alpha + h) - \lambda(\beta).$$

Since however both intervals $(\alpha, \beta]$ and $(\alpha + h, \beta]$ contain the same number of roots of $\omega_n(x)$, it remains only to show that $\lambda(\alpha) = \lambda(\alpha + h)$.

Suppose that the polynomials $\omega_{k_1}(x), \omega_{k_2}(x), \ldots, \omega_{k_r}(x)$ vanish at the point α but that $\omega_n(x)$ does not.

To calculate $\lambda(\alpha)$ we delete all the zeros in the sequence

$$\omega_0(\alpha),\ \omega_1(\alpha),\ \omega_2(\alpha),\ \ldots,\ \omega_n(\alpha),$$

form all products of adjacent numbers in the remaining sequence, and determine the number of negative products among these. In particular, all products

$$\omega_{k_s-1}(\alpha)\omega_{k_s+1}(\alpha) \qquad (s = 1, 2, \ldots, r).$$

are negative.

Proceeding to the point $x = \alpha + h$, each of these products becomes replaced by the two products

$$\omega_{k_s-1}(\alpha + h)\omega_{k_s}(\alpha + h),\quad \omega_{k_s}(\alpha + h)\omega_{k_s+1}(\alpha + h)$$

of which one and only one is negative. All other products $\omega_j(x)\omega_{j+1}(x)$ have at $x = \alpha + h$ the same sign as at $x = \alpha$. This shows that $\lambda(\alpha) = \lambda(\alpha + h)$.

If now $\omega_n(\alpha) = 0$, then in calculating $\lambda(\alpha)$ the product $\omega_{n-1}(\alpha)\omega_n(\alpha)$ is deleted, while it appears when $\lambda(\alpha + h)$ is calculated; since however it is positive, it contributes nothing to $\lambda(\alpha + h)$. Therefore $\lambda(\alpha) = \lambda(\alpha + h)$ holds in this case also.

If β rather than α is a root of some of the polynomials (77), we proceed analogously, i.e. we consider the difference $\lambda(\alpha) - \lambda(\beta + h)$ which is equal to the number of roots of $\omega_n(x)$ in $[\alpha, \beta]$, and we then show that $\lambda(\beta) = \lambda(\beta + h)$.

Finally, the last case in which α and β are both roots of polynomials (77) is dealt with by introducing a point $\gamma(\alpha < \gamma < \beta)$ at which no polynomial (77) vanishes. One then obtains the number of roots of $\omega_n(x)$ in $(\alpha, \beta]$ by addition of the differences $\lambda(\alpha) - \lambda(\gamma)$ and $\lambda(\gamma) - \lambda(\beta)$. This completes the proof.

§3. Relation to the Theory of Continued Fractions

If x is held fixed the fraction

$$\frac{\bar{\omega}_n(t) - \bar{\omega}_n(x)}{t - x} \qquad (n > 0)$$

is a polynomial of degree $n - 1$ in t; but because of the symmetry it is also a polynomial of degree $n - 1$ in x for fixed t. The function

$$\psi_n(x) = \int_a^b p(t)\,\frac{\bar{\omega}_n(t) - \bar{\omega}_n(x)}{t - x}\,dt \qquad (n > 0) \quad (80)$$

is therefore also a polynomial of degree $n - 1$.

Theorem 1. *The recursion formula*

$$\psi_{n+2}(x) = (x - \alpha_{n+2})\psi_{n+1}(x) - \lambda_{n+1}\psi_n(x) \qquad (n = 1, 2, \ldots), \quad (81)$$

holds where α_{n+2} and λ_{n+1} are the same numbers that appear in formula (68).

Proof. From (68)

$$\tilde{\omega}_{n+2}(t) - \tilde{\omega}_{n+2}(x) = t\tilde{\omega}_{n+1}(t) - x\tilde{\omega}_{n+1}(x) - \alpha_{n+2}[\tilde{\omega}_{n+1}(t) - \tilde{\omega}_{n+1}(x)] - \lambda_{n+1}[\tilde{\omega}_n(t) - \tilde{\omega}_n(x)].$$

Whence it follows that

$$\tilde{\omega}_{n+2}(t) - \tilde{\omega}_{n+2}(x) = (t - x)\tilde{\omega}_{n+1}(t) + (x - \alpha_{n+2})[\tilde{\omega}_{n+1}(t) - \tilde{\omega}_{n+1}(x)] - \lambda_{n+1}[\tilde{\omega}_n(t) - \tilde{\omega}_n(x)].$$

Dividing this equation by $(t - x)$, multiplying the result by $p(t)$, and then integrating, we obtain (81) immediately due to

$$\int_a^b p(t)\tilde{\omega}_{n+1}(t)\,dt = 0;$$

this formula also holds for $n = 0$ if we define

$$\psi_0(x) = 0.$$

We now introduce the notation

$$\lambda_0 = \int_a^b p(x)\,dx = \psi_1(x),$$

and α_1 shall denote the root of the polynomial $\omega_1(x)$. The fraction

$$\frac{\psi_n(x)}{\tilde{\omega}_n(x)} \tag{82}$$

is then the nth convergent of the continued fraction

$$\cfrac{\lambda_0}{x - \alpha_1 - \cfrac{\lambda_1}{x - \alpha_2 - \cfrac{\lambda_2}{x - \alpha_3 - \cdots}}}. \tag{83}$$

For the first convergent of (83) is

$$\frac{\lambda_0}{x - \alpha_1} = \frac{\psi_1(x)}{\tilde{\omega}_1(x)},$$

and the second is

$$\frac{\lambda_0}{x - \alpha_1 - \dfrac{\lambda_1}{x - \alpha_2}} = \frac{\psi_1(x)}{\tilde{\omega}_1(x) - \dfrac{\lambda_1}{x - \alpha_2}} = \frac{(x - \alpha_2)\psi_1(x) - \lambda_1\psi_0(x)}{(x - \alpha_2)\tilde{\omega}_1(x) - \lambda_1\tilde{\omega}_0(x)} = \frac{\psi_2(x)}{\tilde{\omega}_2(x)},$$

whence the proof of our theorem follows from the general theorems on continued fractions.

Theorem 2. (T. J. STIELTJES). *If x is a real member not contained in the interval $[a, b]$, then the continued fraction* (83) *converges to the limit*

$$\int_a^b \frac{p(t)\,dt}{x-t}. \tag{84}$$

Proof.[6] Without loss of generality assume $x > b$. We then consider the following function of the n variables $z_1, z_2, \ldots, z_n$:

$$\Phi_n(z_1, z_2, \ldots, z_n) = \int_a^b [1 + z_1(x-t) + z_2(x-t)^2 + \cdots + z_n(x-t)^n]^2 \frac{p(t)\,dt}{x-t}.$$

We then ask for the set of values of $z_1, z_2, \ldots, z_n$ for which this function assumes a minimum.

Such a set of values exists and is even unique, for it is simply a question of finding the best mean approximation to the function $f(t) = -1$ with the weight function $\dfrac{p(t)}{x-t}$ whereby the approximation is made by means of the *linearly independent* functions $x-t, (x-t)^2, \ldots, (x-t)^n$.

As we know, the desired function must satisfy the equations

$$\frac{\partial \Phi_n}{\partial z_1} = 0, \quad \frac{\partial \Phi_n}{\partial z_2} = 0, \quad \ldots, \quad \frac{\partial \Phi_n}{\partial z_n} = 0.$$

If we denote the solution set by $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_n$, then

$$\frac{1}{2}\frac{\partial \Phi_n}{\partial z_i} = \int_a^b p(t)[1 + \bar{z}_1(x-t) + \cdots + \bar{z}_n(x-t)^n](x-t)^{i-1}\,dt = 0 \tag{85}$$

$$(i = 1, 2, \ldots, n).$$

For ease of notation we set

$$1 + \bar{z}_1(x-t) + \bar{z}_2(x-t)^2 + \cdots + \bar{z}_n(x-t)^n = H(t).$$

Then (85) becomes

$$\int_a^b p(t)H(t)(x-t)^{i-1}\,dt = 0 \qquad (i = 1, 2, \ldots, n).$$

Putting i equal to $1, 2, \ldots, n$ in succession, it follows easily that

$$\int_a^b p(t)H(t)t^k\,dt = 0 \qquad (k = 0, 1, \ldots, n-1);$$

[6] STIELTJES [1] (Letter No. 167). Another proof (due to the same author) is to be found in § 3 of Chapter V in the third part of this book.

$H(t)$ is therefore a polynomial of degree n which is orthogonal with weight function $p(t)$ to all lower powers of t. This is however possible only if $H(t)$ differs from $\bar{\omega}_n(t)$ only by a constant factor; therefore

$$H(t) = C\bar{\omega}_n(t).$$

Putting $t = x$, we obtain $1 = C\bar{\omega}_n(x)$, whence it follows that

$$H(t) = \frac{\bar{\omega}_n(t)}{\bar{\omega}_n(x)}.$$

The minimum value of the function Φ_n, which we denote by M_n, is therefore

$$M_n = \int_a^b p(t)\left[\frac{\bar{\omega}_n(t)}{\bar{\omega}_n(x)}\right]^2 \frac{dt}{x-t} = \int_a^b p(t)\frac{\bar{\omega}_n(t)}{\bar{\omega}_n(x)}\left[1 + \sum_{i=1}^{n} \bar{z}_i(x-t)^i\right]\frac{dt}{x-t}.$$

But

$$\int_a^b p(t)\bar{\omega}_n(t)(x-t)^{i-1}\,dt = 0 \qquad (i = 1, 2, \ldots, n),$$

and therefore

$$M_n = \int_a^b p(t)\frac{\bar{\omega}_n(t)}{\bar{\omega}_n(x)}\frac{dt}{x-t} = \int_a^b p(t)\frac{\bar{\omega}_n(t) - \bar{\omega}_n(x) + \bar{\omega}_n(x)}{(x-t)\bar{\omega}_n(x)}\,dt.$$

Whence it follows that

$$M_n = \int_a^b \frac{p(t)\,dt}{x-t} - \frac{\psi_n(x)}{\bar{\omega}_n(x)}.$$

The proof therefore reduces to showing

$$\lim_{n\to\infty} M_n = 0. \tag{86}$$

With this in mind let us consider

$$M_{n-1} = \min\{\Phi_{n-1}(z_1, z_2, \ldots, z_{n-1})\}.$$

Denoting by $z_1^*, z_2^*, \ldots, z_{n-1}^*$ the set of values for which Φ_{n-1} assumes a minimum, then clearly for every λ

$$0 \leqq M_n \leqq \int_a^b p(t)\{[1 + z_1^*(x-t) + \cdots + z_{n-1}^*(x-t)^{n-1}][1 - \lambda(x-t)]\}^2\frac{dt}{x-t},$$

since the integral on the right-hand side is some value of Φ_n.

We now put $\lambda = \dfrac{1}{x-a}$. Taking into account the estimate

$$0 \leqq 1 - \frac{x-t}{x-a} \leqq \frac{b-a}{x-a} = q < 1$$

which is valid for $a \leqq t \leqq b$, we find

$$0 \leqq M_n \leqq q^2 \int_a^b p(t)[1 + z_1^*(x-t) + \cdots + z_{n-1}^*(x-t)^{n-1}]^2 \frac{dt}{x-t};$$

This means however that

$$0 \leqq M_n \leqq q^2 M_{n-1}.$$

Since analogously

$$0 \leqq M_{n-1} \leqq q^2 M_{n-2},$$

$$\cdots\cdots\cdots\cdots$$

$$0 \leqq M_2 \leqq q^2 M_1,$$

it follows that

$$0 \leqq M_n \leqq q^{2n-2} M_1,$$

whence we obtain (86). We further note that convergence of the fractions $\frac{\psi_n(x)}{\tilde{\omega}_n(x)}$ to the integral (84) is uniform on any set which is a positive distance from the interval $[a, b]$.

The integral (84) has another interesting expansion. For $|t| < |x|$

$$\frac{1}{1 - \frac{t}{x}} = \sum_{k=0}^{\infty} \frac{t^k}{x^k},$$

where convergence of the series is uniform for $t \in [a, b]$ if [7]

$$|x| > \max\{|a|, |b|\}$$

For such values of x then

$$\int_a^b \frac{p(t)}{x-t}\,dt = \frac{1}{x}\int_a^b \frac{p(t)}{1-\frac{t}{x}}\,dt = \frac{1}{x}\int_a^b p(t)\left[\sum_{k=0}^{\infty}\frac{t^k}{x^k}\right]dt.$$

Because of the uniform convergence we may integrate termwise; this gives

$$\int_a^b \frac{p(t)}{x-t}\,dt = \sum_{k=0}^{\infty} \frac{1}{x^{k+1}} \int_a^b p(t)t^k\,dt. \tag{87}$$

[7] For if $C = \max\{|a|, |b|\}$, then $[a, b] \subset [-C, C]$ and $t \in [a, b]$ imply $|t| \leqq C$.

The integrals on the right are the moments of the weight function $p(t)$, so that we finally obtain

$$\int_a^b \frac{p(t)}{x-t}\,dt = \sum_{k=0}^{\infty} \frac{\mu_k}{x^{k+1}}.$$

Theorem 3. *The convergent* $\dfrac{\psi_n(x)}{\bar{\omega}_n(x)}$ *has the property that*

$$\lim_{n\to+\infty} x^{2n+1}\left[\int_a^b \frac{p(t)}{x-t}\,dt - \frac{\psi_n(x)}{\bar{\omega}_n(x)}\right] \tag{88}$$

exists and is finite; moreover, it is the only rational function whose denominator is at most of degree n with this property.

Proof. First, the expression for $\psi_n(x)$ gives

$$\int_a^b \frac{p(t)}{x-t}\,dt - \frac{\psi_n(x)}{\bar{\omega}_n(x)} = \frac{1}{\bar{\omega}_n(x)}\int_a^b p(t)\,\frac{\bar{\omega}_n(t)}{x-t}\,dt.$$

We here again convince ourselves, as in equation (87), that for sufficiently large values of x

$$\int_a^b \frac{p(t)\bar{\omega}_n(t)}{x-t}\,dt = \sum_{k=0}^{\infty} \frac{1}{x^{k+1}}\int_a^b p(t)\bar{\omega}_n(t)t^k\,dt.$$

But now for $k < n$ the functions t^k and $\bar{\omega}_n(t)$ are orthogonal with weight function $p(t)$, whence

$$\int_a^b \frac{p(t)}{x-t}\,dt - \frac{\psi_n(x)}{\bar{\omega}_n(x)} = \frac{1}{\bar{\omega}_n(x)}\left[\frac{c_n}{x^{n+1}} + \frac{c_{n+1}}{x^{n+2}} + \cdots\right].$$

Since now

$$\lim_{n\to+\infty} \frac{x^n}{\bar{\omega}_n(x)} = 1,$$

the limit (88) exists and is equal to c_n.

Noting that

$$t^n = \bar{\omega}_n(t) + \rho(t)$$

where the degree of the polynomial $\rho(t)$ is less than n, it follows that

$$c_n = \int_a^b p(t)\bar{\omega}_n(t)t^n\,dt = \int_a^b p(t)\bar{\omega}_n^2(t)\,dt = \frac{\Delta_n}{\Delta_{n-1}}.$$

We must now prove that there are no other rational functions with the stated properties. Suppose

$$\frac{h(x)}{q(x)} \qquad (q(x) = a_0x^n + a_1x^{n-1} + \cdots)$$

is such a function; then the limit

$$\lim_{x\to+\infty} x^{2n+1}\left[\frac{h(x)}{q(x)} - \frac{\psi_n(x)}{\tilde{\omega}_n(x)}\right]$$

must exist.

On the other hand, the finite limit

$$\lim_{x\to+\infty} \frac{q(x)\tilde{\omega}_n(x)}{x^{2n}} = a_0$$

clearly exists. Therefore

$$\lim_{x\to+\infty} x[h(x)\tilde{\omega}_n(x) - q(x)\psi_n(x)]$$

exists and is finite which is only possible if

$$h(x)\tilde{\omega}_n(x) - q(x)\psi_n(x) = 0.$$

Therefore $\frac{h(x)}{q(x)}$ coincides with the convergent $\frac{\psi_n(x)}{\tilde{\omega}_n(x)}$.

We note further that the rational function $\frac{\psi_n(x)}{\tilde{\omega}_n(x)}$ is irreducible, for from the recursion formulas (68) and (81)

$$\psi_{n+1}(x) = (x - \alpha_{n+1})\psi_n(x) - \lambda_n\psi_{n-1}(x);$$
$$\tilde{\omega}_{n+1}(x) = (x - \alpha_{n+1})\tilde{\omega}_n(x) - \lambda_n\tilde{\omega}_{n-1}(x),$$

whence it follows that

$$\psi_{n+1}(x)\tilde{\omega}_n(x) - \psi_n(x)\tilde{\omega}_{n+1}(x) = \lambda_n[\psi_n(x)\tilde{\omega}_{n-1}(x) - \psi_{n-1}(x)\tilde{\omega}_n(x)].$$

Continuing the procedure on the right-hand side, we arrive at the formula

$$\psi_{n+1}(x)\tilde{\omega}_n(x) - \psi_n(x)\tilde{\omega}_{n+1}(x) = \lambda_n\lambda_{n-1}\ldots\lambda_1\lambda_0 > 0, \tag{89}$$

which is itself of interest. It implies in particular that $\psi_n(x)$ and $\tilde{\omega}_n(x)$ have no common roots.

We can draw still another conclusion from Theorem 3 and the irreducibility of the rational function $\frac{\psi_n(x)}{\tilde{\omega}_n(x)}$. If the denominator of the rational function $\frac{h(x)}{q(x)}$ is of degree at most n and if a finite limit

$$\lim_{x \to +\infty} x^{2n+1} \left[\int_a^b \frac{p(t)}{x - t}\, dt - \frac{h(x)}{q(x)} \right],$$

exists, then $q(x)$ and $\bar{\omega}_n(x)$ differ at most by a constant factor.[8] This property forms the basis for a method of constructing the polynomials $\bar{\omega}_n(x)$; it does not appear very practical to us however.

We shall now illustrate the theory presented by an example, namely the TCHEBYSHEFF polynomials, which as we know from the first part of the book form an orthogonal system on the interval $[-1, 1]$ with the weight $\frac{1}{\sqrt{1 - x^2}}$.

The recursion formula for the TCHEBYSHEFF polynomials reads

$$T_{n+2}(x) = 2xT_{n+1}(x) - T_n(x),$$

but for $n > 0$ the leading coefficient of $T_n(x)$ is 2^{n-1}. For the polynomials $\tilde{T}_n(x) = \frac{1}{2^{n-1}} T(x)$; therefore

$$\tilde{T}_{n+2}(x) = x\tilde{T}_{n+1}(x) - \tfrac{1}{4}\tilde{T}_n(x). \qquad (n = 1, 2, \ldots)$$

Hence, the numbers denoted by α_{n+2} and λ_{n+1} in the general theory are

$$\alpha_{n+2} = 0, \quad \lambda_{n+1} = \tfrac{1}{4}. \qquad (n = 1, 2, \ldots)$$

If $n = 0$, then

$$T_2(x) = 2xT_1(x) - T_0(x),$$

and

$$T_0(x) = 1 = \tilde{T}_0(x).$$

Therefore

$$\tilde{T}_2(x) = xT_1(x) - \tfrac{1}{2}\tilde{T}_0(x),$$

and hence

$$\alpha_2 = 0 \quad \text{and} \quad \lambda_1 = \tfrac{1}{2}.$$

Moreover, α_1 is the root of the polynomial $T_1(x) = x$, and hence $\alpha_1 = 0$. Finally,

$$\lambda_0 = \int_{-1}^{+1} \frac{dx}{\sqrt{1 - x_2}} = \pi,$$

and hence

[8] For $\frac{h(x)}{q(x)} = \frac{\psi_n(x)}{\bar{\omega}_n(x)}$, and hence $\frac{q(x)\psi_n(x)}{\bar{\omega}_n(x)} = h(x)$, so that $\bar{\omega}_n(x)$ is a factor of $q(x)$.

$$\int_{-1}^{+1} \frac{dt}{(x-t)\sqrt{1-t^2}} = \cfrac{\pi}{x - \cfrac{\frac{1}{2}}{x - \cfrac{\frac{1}{4}}{x - \cdots}}}.$$

Using the substitution $t = \dfrac{2u}{1+u^2}$, the last integral is easily computed:

$$\int_{-1}^{+1} \frac{dt}{(x-t)\sqrt{1-t^2}} = \frac{\pi}{\sqrt{x^2-1}} \qquad (x > 1),$$

whence we obtain

$$\frac{1}{\sqrt{x^2-1}} = \cfrac{1}{x - \cfrac{\frac{1}{2}}{x - \cfrac{\frac{1}{4}}{x - \cfrac{\frac{1}{4}}{x - \cdots}}}} \qquad (x > 1).$$

The denominators of the continued fractions therefore coincide, as they must, with the polynomials $\tilde{T}_n(x)$.

From the binomial formula of NEWTON

$$\frac{\pi}{\sqrt{x^2-1}} = \frac{\pi}{x}\left(1 - \frac{1}{x^2}\right)^{-1/2} = \frac{\pi}{x}\left[1 + \frac{1}{2}\,\frac{1}{x^2} + \frac{3!!}{2^2 \cdot 2!}\,\frac{1}{x^4} + \frac{5!!}{2^2 \cdot 3!}\,\frac{1}{x^6} + \cdots\right].$$

The moments of the weight function $\dfrac{1}{\sqrt{1-x^2}}$ are therefore

$$\mu_0 = \int_{-1}^{+1} \frac{dx}{\sqrt{1-x^2}} = \pi,$$

$$\mu_{2n-1} = \int_{-1}^{+1} \frac{x^{2n-1}\,dx}{\sqrt{1-x^2}} = 0,$$

$$\mu_{2n} = \int_{-1}^{+1} \frac{x^{2n}\,dx}{\sqrt{1-x^2}} = \frac{(2n-1)!!}{2^n n!}\,\pi.$$

§4. The CHRISTOFFEL-DARBOUX Formula. Convergence of Orthogonal Expansions

Let $\{\omega_n(x)\}$ be an orthogonal system with weight $p(x)$, and let $f(x)$ be any function of $L_{p(x)}$.

The FOURIER series for this function is

$$\sum_{k=0}^{\infty} c_k \omega_k(x) \quad \left(c_k = \int_a^b p(t) f(t) \omega_k(t)\, dt \right). \tag{90}$$

Denoting the partial sum

$$\sum_{k=0}^{n} c_k \omega_k(x)$$

by $S_n(x)$, we obtain

$$S_n(x) = \int_a^b p(t) f(t) \left[\sum_{k=0}^{n} \omega_k(t) \omega_k(x) \right] dt. \tag{91}$$

The expression

$$K_n(t, x) = \sum_{k=0}^{n} \omega_k(t) \omega_k(x) \tag{92}$$

is called the *kernel* of the integral (91). It plays a major role in the study of the convergence of the series (90).

Theorem 1. *The following formula holds:*

$$K_n(t, x) = \sqrt{\lambda_{n+1}} \, \frac{\omega_{n+1}(t)\omega_n(x) - \omega_{n+1}(x)\omega_n(t)}{t - x}. \tag{93}$$

This is called the CHRISTOFFEL-DARBOUX formula.[9]

Proof. For the proof we use the recursion formula (70) in the form:

$$\sqrt{\frac{\Delta_{n+1}}{\Delta_n}} \, \omega_{n+1}(x) = (x - \alpha_{n+1}) \sqrt{\frac{\Delta_n}{\Delta_{n-1}}} \, \omega_n(x) - \lambda_n \sqrt{\frac{\Delta_{n-1}}{\Delta_{n-2}}} \, \omega_{n-1}(x) \quad (n \geqq 1). \tag{94}$$

Here

$$\lambda_n = \frac{\Delta_{n-2}\Delta_n}{\Delta_{n-1}^2}. \tag{95}$$

We multiply (94) by $\omega_n(t)$ and subtract the result from the equation obtained by interchanging t and x:

$$\sqrt{\frac{\Delta_{n+1}}{\Delta_n}} [\omega_{n+1}(x)\omega_n(t) - \omega_{n+1}(t)\omega_n(x)] = \sqrt{\frac{\Delta_n}{\Delta_{n-1}}} (x - t)\omega_n(t)\omega_n(x) + \lambda_n \sqrt{\frac{\Delta_{n-1}}{\Delta_{n-2}}} [\omega_n(x)\omega_{n-1}(t) - \omega_n(t)\omega_{n-1}(x)].$$

[9] CHRISTOFFEL [1], p. 73 proved it for $p(x) = 1$, $a = -1$, $b = +1$, i.e. for the LEGENDRE polynomials. DARBOUX [1], p. 411 generalized it to arbitrary weight functions.

Multiplying this equation by $\sqrt{\frac{\Delta_{n-1}}{\Delta_n}}$ and using (95), we obtain

$$\sqrt{\lambda_{n+1}}\,[\omega_{n+1}(x)\omega_n(t) - \omega_{n+1}(t)\omega_n(x)] = (x-t)\omega_n(t)\omega_n(x) + \sqrt{\lambda_n}\,[\omega_n(x)\omega_{n-1}(t) - \omega_n(t)\omega_{n-1}(x)]. \quad (96)$$

If we here substitute the values n, $n-1$, $n-2$, ..., 1 successively in place of n and add the n equations so obtained, after simple rearrangements we obtain the equation

$$\sqrt{\lambda_{n+1}}\,[\omega_{n+1}(x)\omega_n(t) - \omega_{n+1}(t)\omega_n(x)] = (x-t)\sum_{k=1}^{n}\omega_k(t)\omega_k(x) + \sqrt{\lambda_1}\,[\omega_1(x)\omega_0(t) - \omega_1(t)\omega_0(x)].$$

But

$$\omega_0(x) = \omega_0(t) = \frac{1}{\sqrt{\Delta_0}}, \quad \omega_1(x) = \sqrt{\frac{\Delta_0}{\Delta_1}}\,(x-\alpha_1), \quad \lambda_1 = \frac{\Delta_1}{\Delta_0^2},$$

and therefore

$$\sqrt{\lambda_1}\,[\omega_1(x)\omega_0(t) - \omega_1(t)\omega_0(x)] = \frac{x-t}{\Delta_0} = (x-t)\omega_0(t)\omega_0(x),$$

whence (93) follows.

As proved above, we have the estimate

$$\sqrt{\lambda_{n+1}} \leqq C = \max\{|a|, |b|\}.$$

Therefore

$$K_n(t,x) = \theta_n C\,\frac{\omega_{n+1}(t)\omega_n(x) - \omega_{n+1}(x)\omega_n(t)}{t-x} \quad (0 < \theta_n \leqq 1).$$

From (91) with $f(t) \equiv 1$ it follows easily that

$$\int_a^b p(t)K_n(t,x)\,dt = 1,$$

so that we obtain

$$S_n(x) - f(x) = \theta_n C\int_a^b p(t)\varphi_x(t)[\omega_{n+1}(t)\omega_n(x) - \omega_{n+1}(x)\omega_n(t)]\,dt \quad (97)$$

where for brevity we have defined

$$\varphi_x(t) = \frac{f(t)-f(x)}{t-x}. \quad (98)$$

Suppose now that $\varphi_x(t) \in L^2_{p(t)}$ (which is possible only if $f(t)$ also belongs to this class).

Denoting the FOURIER coefficients of the function $\varphi_x(t)$ by d_n, we can write equation (97) in the form

$$S_n(x) - f(x) = \theta_n C[d_{n+1}\omega_n(x) - d_n\omega_{n+1}(x)], \tag{99}$$

and since the series $\sum d_n^2$ converges

$$\lim_{n\to\infty} d_n = 0.$$

This and (99) give

Theorem 2. *If all the polynomials $\omega_n(x)$ are uniformly bounded with respect to n at the point x and if the function $\varphi_x(t)$ belongs to $L^2_{p(t)}$, then at the point x the following equation holds:*

$$f(x) = \sum_{k=0}^{\infty} c_k\omega_k(x).$$

If the polynomials $\omega_n(x)$ are uniformly bounded on the entire domain of definition $[a, b]$ (the TCHEBYSHEFF polynomials for example), then the condition $\varphi_x(t) \in L^2_{p(t)}$ in Theorem 2 can be relaxed, and instead of this it suffices that $\varphi_x(t) \in L_{p(t)}$.

To prove this one needs

Theorem 3. *If the polynomials $\omega_n(x)$ are uniformly bounded (with respect to both n and x), then for every function $\varphi(x)$ of $L_{p(x)}$*

$$\lim_{n=\infty} \int_a^b p(x)\varphi(x)\omega_n(x)\,dx = 0.$$

Proof. Because of the absolute continuity of the integral, for any given $\varepsilon > 0$, it is possible to find a number $\delta > 0$ such that

$$\int_a p(x)|\varphi(x)|\,dx < \varepsilon$$

for any set $e \subset [a, b]$ whose measure $me < \delta$. On the other hand, it is possible to find a number K so large that

$$mE(|\varphi| > K) < \delta.$$

We now define the functions

$$\varphi_1(x) = \begin{cases} \varphi(x), & \text{for } |\varphi(x)| \leqq K, \\ 0, & \text{for } |\varphi(x)| > K. \end{cases}$$

$$\varphi_2(x) = \begin{cases} 0, & \text{for } |\varphi(x)| \leqq K, \\ \varphi(x), & \text{for } |\varphi(x)| > K. \end{cases}$$

Then clearly

$$\int_a^b p\varphi\omega_n\,dx = \int_a^b p\varphi_1\omega_n\,dx + \int_a^b p\varphi_2\omega_n\,dx.$$

The function $\varphi_1(x)$ is bounded and hence belongs to $L^2_{p(x)}$, and its FOURIER coefficients tend to zero, such that for sufficiently large n the inequality

$$\left| \int_a^b p(x)\varphi_1(x)\omega_n(x)\,dx \right| < \varepsilon$$

holds.

If on the other hand M is a bound for $|\omega_n(x)|$, then for all n

$$\left| \int_a^b p(x)\varphi_2(x)\omega_n(x)\,dx \right| \leqq M \int_{E(|\varphi|>K)}^b p(x)|\varphi(x)|\,dx < M\varepsilon,$$

and hence for large values of n:

$$\left| \int_a^b p(x)\varphi(x)\omega_n(x)\,dx \right| < (M+1)\varepsilon,$$

which completes the proof. From this theorem and (99) we obtain

Theorem 4. *If there exists a uniformly bounded orthonormal system of polynomials on* $[a, b]$, *then every function* $f(x) \in L_{p(x)}$ *for which* $\varphi_x(t)$ *also belongs to* $L_{p(t)}$ *can be developed in the corresponding* FOURIER *series at every point* $x \in [a, b]$.

In particular, this theorem is applicable to functions belonging to Lip α ($\alpha > 0$ arbitrary) if in addition the weight $p(x)$ is *bounded* and to functions belonging to the class Lip 1 for any arbitrary weight function.

The results obtained lead to the following theorem which expresses the so-called "localization principle":

Theorem 5. *If two functions* $f(x)$ *and* $g(x)$ *belonging to* $L^2_{p(x)}$ *coincide on the interval* $(x_0 - h, x_0 + h)$ *and if the polynomials* $\omega_n(x)$ *are uniformly bounded with respect to* n *at the point* x_0, *then*

$$\lim_{n\to\infty} \{S_n[f; x_0] - S_n[g; x_0]\} = 0.$$

Here $S_n[f; x_0]$ denotes the partial FOURIER sum of the function $f(x)$ at the point x_0.

Proof. For the proof we note that

$$S_n[f; x_0] - S_n[g; x_0] = S_n[r; x_0],$$

where $r(x) = f(x) - g(x)$. Since $r(x_0) = 0$ the proof follows from Theorem 2, because the function $\dfrac{r(t)}{x_0 - t}$ vanishes on the interval $(x_0 - h, x_0 + h)$ and therefore belongs to $L^2_{p(t)}$.

If the polynomials $\omega_n(x)$ are in addition *uniformly* bounded with respect

to x on the entire interval $[a, b]$, then according to Theorem 3 it suffices that $f(x)$ and $g(x)$ belong to $L_{p(x)}$.

If the condition that the polynomials $\omega_n(x)$ be bounded is dropped the preceding theorems can no longer be applied. Nevertheless, we still have at least

Theorem 6. (I. P. NATANSON [1]). *If $f(x) \in Lip\ \alpha$ and $\alpha > \frac{1}{2}$, then*

$$f(x) = \sum_{k=0}^{\infty} c_k \omega_k(x) \tag{100}$$

almost everywhere on a, b.

Proof. From the JACKSON theorem of Vol. 1, Ch. VI, § 2, for every n there exists a polynomial $P_n(x) \in H_n$ such that

$$|P_n(x) - f(x)| < \frac{A}{n^\alpha},$$

where A is a constant. But from the TOEPLER theorem the partial FOURIER sum $S_n(x)$ satisfies the condition

$$\int_a^b p(x)[S_n(x) - f(x)]^2\, dx \leqq \int_a^b p(x)[P_n(x) - f(x)]^2\, dx,$$

so that

$$\int_a^b p(x)[S_n(x) - f(x)]^2\, dx < \frac{A^2 \mu_0}{n^{2\alpha}} \qquad \left(\mu_0 = \int_a^b p(x)\, dx\right),$$

and therefore the series

$$\sum_{n=1}^{\infty} \int_a^b p(x)[S_n(x) - f(x)]^2\, dx$$

converges. The assertion now follows from the LEVI theorem.

With more powerful tools A. N. KOLMOGOROFF [2] was able to extend this result to $\alpha > 0$, and using these same more powerful means I. P. NATANSON [2] proved that any function of bounded variation can be expanded in the polynomials $\omega_n(x)$ almost everywhere if the weight $p(x)$ satisfies the condition

$$p(x) < \frac{A}{\sqrt{1 - x^2}}.$$

All of these investigations have one shortcoming in common: they lack a characterization of those points where the expansion (100) is valid.

There is another approach to the problems arising in the expansion of a function in terms of orthogonal polynomials which relates them to the theory

of uniform approximation. The key to this approach is the fact that for every polynomial $P_n(x) \in H_n$ the equation

$$S_n[P_n; x] = P_n(x)$$

holds. We define

$$\int_a^b p(t)|K_n(t, x)|\, dt = L_n(x)$$

and call $L_n(x)$ a LEBESGUE *function* of the orthonormal system $\{\omega_n(x)\}$. If $f(x)$ is bounded, $|f(x)| \leqq M$, then formula (91) yields the estimate

$$|S_n[f; x_0]| \leqq ML_n(x_0).$$

From this follows

Theorem 7. *Suppose that the continuous function $f(x)$ has the property that its best approximation by polynomials $E_n(f)$ satisfies the condition*

$$L_n(x_0)E_n(f) \underset{n\to\infty}{\longrightarrow} 0,$$

then the function can be expanded into a FOURIER *series with the polynomials $\omega_n(x)$ at the point x_0.*

Proof. If $P_n(x)$ is the polynomial of best approximation to $f(x)$, then

$$|P_n(x) - f(x)| \leqq E_n(f).$$

Hence

$$f(x) - S_n[f; x] = f(x) - P_n(x) + S_n[P_n - f; x],$$

and therefore

$$|f(x_0) - S_n[f; x_0]| \leqq E_n(f) + L_n(x_0)E_n(f), \tag{101}$$

which completes the proof of the theorem.

We now define

$$L_n = \max L_n(x).$$

If the condition

$$\lim_{n\to\infty} L_nE_n(f) = 0$$

is satisfied we can—as the estimate (101) shows—guarantee the uniform convergence of the series

$$\sum_{k=0}^{\infty} c_k\omega_k(x) = f(x).$$

Example. We shall find the LEBESGUE functions for the orthonormal system of TCHEBYSHEFF polynomials [10]

[10] Following V. L. GONCHAROV we denote the *normalized* TCHEBYSHEFF polynomials by $T_n(x)$.

$$\hat{T}_0(x) = \frac{1}{\sqrt{\pi}}, \qquad \hat{T}_n(x) = \sqrt{\frac{2}{\pi}} \cos (n \arccos x)$$

Here

$$L_n(x) = \int_{-1}^{+1} \left| \sum_{k=0}^{n} \hat{T}_k(t)\hat{T}_k(x) \right| \frac{dt}{\sqrt{1-t^2}}.$$

With $\arccos x = \xi$ and $\arccos t = \tau$ we find

$$L_n(x) = \frac{1}{\pi} \int_0^{\pi} \left| 1 + 2 \sum_{k=1}^{n} \cos k\tau \cos k\xi \right| d\tau,$$

whence, since the integrand is an even function,

$$L_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 + \sum_{k=1}^{n} [\cos k(\tau + \xi) + \cos k(\tau - \xi)] \right| d\tau,$$

and from this

$$L_n(x) \leqq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{2} + \sum_{k=1}^{n} \cos k(\tau + \xi) \right| d\tau + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{2} + \sum_{k=1}^{n} \cos k(\tau - \xi) \right| d\tau.$$

If we substitute $\tau = \varphi - \xi$ in the first integral and $\tau = \varphi + \xi$ in the second integral (where in so doing we may retain the same limits of integration due to the 2π-periodicity) we obtain

$$L_n(x) \leqq \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \frac{1}{2} + \sum_{k=1}^{n} \cos k\varphi \right| d\varphi.$$

Now from formula (175) of Vol. 1

$$\frac{1}{2} + \sum_{k=1}^{n} \cos k\varphi = \frac{\sin \dfrac{2n+1}{2} \varphi}{2 \sin \dfrac{\varphi}{2}},$$

whence

$$L_n(x) \leqq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\sin \dfrac{2n+1}{2} \varphi}{\sin \dfrac{\varphi}{2}} \right| d\varphi$$

or

$$L_n(x) \leqq \frac{2}{\pi} \int_0^{\pi/2} \left| \frac{\sin (2n+1)t}{\sin t} \right| dt,$$

from which using estimate (178) of Vol. 1 we obtain finally

$$L_n(x) \leqq 2 + \ln n.$$

Therefore, all functions for which

$$\lim_{n \to \infty} [E_n(f) \ln n] = 0$$

can be expanded in a uniformly convergent series of TCHEBYSHEFF polynomials.

This last condition is according to the JACKSON theorem automatically satisfied if the modulus of continuity of the function $f(x)$ fulfills the relation

$$\lim_{\delta \to 0} [\omega(\delta) \ln \delta] = 0,$$

whence we have for the second time obtained the theorem of § 1, Chapter X in Vol. I of this book.

We give finally an especially strong negative assertion which complements the considerations of this section.

Theorem 8. (V. F. NIKOLAEV [1]). *There is no weight function $p(x)$ such that every continuous function can be expanded in a* FOURIER *series of orthogonal functions (with this weight).*

§5. Changing the Weight Function

As we have seen, the boundedness of the polynomials $\omega_n(x)$ at particular points and on the entire interval $[a, b]$ plays an important role in questions of the convergence of orthogonal expansions. We are therefore interested in the question of what conditions we must impose on the weight function in order that such boundedness ensue. For certain weights, e.g. the TCHEBYSHEFF weight $(1 - x)^{-\frac{1}{2}}$, we already know that the corresponding polynomials are bounded. We are thus in a natural way led to the

Problem. *Suppose that an orthonormal system of polynomials $\omega_n(x)$ with weight $p(x)$ is bounded at the point x_0:*

$$|\omega_n(x_0)| \leqq M \qquad (n = 0, 1, 2, \ldots).$$

Now let us consider a new weight $p(x)\sigma(x)$[11] *and the orthonormal system corresponding to it. We now ask what property of the factor $\sigma(x)$ will ensure the boundedness of the system $\{\varphi_n(x)\}$ at the point x_0.*

[11] Of course $\sigma(x)$ is non-negative, belongs to $L_{p(x)}$, and is nonzero almost everywhere.

Theorem 1. *If $\sigma(x)$ is a polynomial, then*

$$|\varphi_n(x_0)| \leqq \frac{M\sqrt{K}\,(m+1)}{\sigma(x_0)}, \tag{102}$$

where K is the maximum and m is the degree of $\sigma(x)$.

Proof. The product $\sigma(x)\varphi_n(x)$ can be represented in the form

$$\sigma(x)\varphi_n(x) = c_0\omega_0(x) + c_1\omega_1(x) + \cdots + c_{n+m}\omega_{n+m}(x), \tag{103}$$

where

$$c_i = \int_a^b p(x)\sigma(x)\varphi_n(x)\omega_i(x)\,dx.$$

If $i < n$, then $c_i = 0$, since $\varphi_n(x)$ is orthogonal with weight $p(x)\sigma(x)$ to all polynomials of lower order. Therefore (103) has only $m+1$ summands. Now suppose $n \leqq i \leqq n+m$. According to the SCHWARZ inequality

$$c_i^2 \leqq \left(\int_a^b p(x)\sigma^2(x)\varphi_n^2(x)\,dx\right)\left(\int_a^b p(x)\omega_i^2(x)\,dx\right).$$

The last integral is equal to one, and further

$$\int_a^b p(x)\sigma^2(x)\varphi_n^2(x)\,dx \leqq K\int_a^b p(x)\sigma(x)\varphi_n^2(x)\,dx = K.$$

Therefore

$$|c_i| \leqq \sqrt{K},$$

which proves the validity of estimate (102).

Corollary 1. *If the system $\{\omega_n(x)\}$ is uniformly bounded on $[a, b]$ and if $\sigma(x)$ has a positive lower bound, then $\{\varphi_n(x)\}$ is also uniformly bounded on $[a, b]$.*

Corollary 2. *The systems of polynomials with weights*

$$(1-x)^{A-\frac{1}{2}}(1+x)^{B-\frac{1}{2}} \qquad (-1 \leqq x \leqq 1), \tag{104}$$

where A and B are natural numbers, are bounded at every point of the interval $(-1, 1)$.

Theorem 2. (J. PEEBLES [1]). *If $\sigma(x) = \dfrac{1}{Q(x)}$ where $Q(x)$ is a polynomial of degree m, then* [12]

$$|\varphi_n(x_0)| \leqq M\sqrt{K}\,(m+1), \tag{105}$$

where $K = \max Q(x)$.

[12] PEEBLES [1], p. 92.

Proof. In the equation

$$\varphi_n(x) = c_0\omega_0(x) + c_1\omega_1(x) + \cdots + c_n\omega_n(x)$$

first of all,

$$c_i = \int_a^b p(x)\varphi_n(x)\omega_i(x)\,dx.$$

For $i < n - m$

$$c_i = \int_a^b p(x)\sigma(x)\varphi_n(x)[Q(x)\omega_i(x)]\,dx = 0$$

and therefore

$$\varphi_n(x) = c_{n-m}\omega_{n-m}(x) + \cdots + c_n\omega_n(x). \tag{106}$$

For $n - m \leqq i \leqq n$ the SCHWARZ inequality gives

$$c_i^2 \leqq \int_a^b p(x)\varphi_n^2(x)\,dx \leqq K\int_a^b p(x)\sigma(x)\varphi_n^2(x)\,dx = K,$$

which together with (106) gives (105).

A combination of Theorems 1 and 2 makes it possible without much difficulty to obtain estimates also in the cases in which $\sigma(x)$ is a fractional rational function.

Theorem 3. (J. KOROUS [1]). *If $\sigma(x)$ satisfies the* LIPSCHITZ *condition*

$$|\sigma(y) - \sigma(x)| \leqq K|y - x|$$

and possesses a positive minimum τ, then

$$|\varphi_n(x_0)| \leqq \frac{\tau + 2CK}{\sqrt{\tau^3}}\,M \qquad (C = \max\{|a|, |b|\}).$$

Proof. We write the polynomial $\varphi_n(x)$ in the form

$$\begin{aligned}\varphi_n(x) &= \int_a^b p(t)\varphi_n(t)K_n(t, x)\,dt \\ &= \int_a^b p(t)\varphi_n(t)\omega_n(t)\omega_n(x)\,dt + \int_a^b p(t)\varphi_n(t)K_{n-1}(t, x)\,dt.\end{aligned} \tag{107}$$

Since $K_{n-1}(t, x)$ is a polynomial of degree $n - 1$ in t,

$$\int_a^b p(t)\sigma(t)\varphi_n(t)K_{n-1}(t, x)\,dt = 0$$

and therefore

$$\sigma(x)\int_a^b p(t)\varphi_n(t)K_{n-1}(t, x)\,dt = \int_a^b p(t)[\sigma(x) - \sigma(t)]\varphi_n(t)K_{n-1}(t, x)\,dt.$$

According to the CHRISTOFFEL-DARBOUX formula

$$K_{n-1}(t, x) = \sqrt{\lambda_n}\,\frac{\omega_n(x)\omega_{n-1}(t) - \omega_n(t)\omega_{n-1}(x)}{x - t}.$$

The last integral of equation (107) may also be written in the form

$$\frac{\sqrt{\lambda_n}}{\sigma(x)}\int_a^b p(t)\,\frac{\sigma(x) - \sigma(t)}{x - t}\,\varphi_n(t)[\omega_n(x)\omega_{n-1}(t) - \omega_n(t)\omega_{n-1}(x)]\,dt.$$

But now

$$\left|\int_a^b p(t)\,\frac{\sigma(x) - \sigma(t)}{x - t}\varphi_n(t)\,\omega_n(t)\,dt\right| \leqq K\int_a^b p(t)|\varphi_n(t)||\omega_n(t)|\,dt,$$

and according to the SCHWARZ inequality

$$\int_a^b p(t)|\varphi_n(t)|\,|\omega_n(t)|\,dt \leqq \frac{1}{\sqrt{\tau}}\int_a^b p(t)\{\sqrt{\sigma(t)}\,|\varphi_n(t)|\}\,|\omega_n(t)|\,dt$$

$$\leqq \frac{1}{\sqrt{\tau}}\left\{\int_a^b p(t)\sigma(t)\varphi_n^2(t)\,dt\right\}^{\frac{1}{2}}\left\{\int_a^b p(t)\omega_n^2(t)\,dt\right\}^{\frac{1}{2}} = \frac{1}{\sqrt{\tau}}, \qquad (108)$$

and this estimate remains valid if we replace $\omega_n(t)$ by $\omega_{n-1}(t)$.

From (107) and (108)

$$|\varphi_n(x)| \leqq \frac{|\omega_n(x)|}{\sqrt{\tau}} + \frac{\sqrt{\lambda_n}}{\sigma(x)}K\,\frac{|\omega_n(x)| + |\omega_{n-1}(x)|}{\sqrt{\tau}},$$

and since $\sqrt{\lambda_n} \leqq C$ the theorem follows.

In the case of Theorem 1 where $\sigma(x)$ is a polynomial the polynomials $\varphi_n(x)$ can be explicitly represented in terms of the polynomials $\omega_n(x)$. Let

$$\sigma(x) = A(x - \xi_1)(x - \xi_2)\ldots(x - \xi_m)$$

be the factorization of $\sigma(x)$ into linear factors; we hereby restrict ourselves to the case in which $\sigma(x)$ has only simple roots. The roots (real as well as the complex conjugate) then lie outside the interval (a, b), since $\sigma(x)$ does not change sign on this interval. We now define

$$Q_n(x) = \begin{vmatrix} \omega_n(\xi_1) & \omega_n(\xi_2) & \cdots & \omega_n(\xi_m) & \omega_n(x) \\ \omega_{n+1}(\xi_1) & \omega_{n+1}(\xi_2) & \cdots & \omega_{n+1}(\xi_m) & \omega_{n+1}(x) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \omega_{n+m}(\xi_1) & \omega_{n+m}(\xi_2) & \cdots & \omega_{n+m}(\xi_m) & \omega_{n+m}(x) \end{vmatrix}. \qquad (109)$$

$Q_n(x)$ is clearly a polynomial of degree $n + m$ which has the points $\xi_1, \xi_2, \ldots, \xi_m$ as roots. The fraction

$$\frac{Q_n(x)}{\sigma(x)}$$

is therefore a polynomial of degree n. On the other hand, for every polynomial $R(x)$ of degree $i < n$ the equation

$$\int_a^b p(x)Q_n(x)R_i(x)\,dx = 0 \tag{110}$$

holds, since $Q_n(x)$ is a linear combination of the polynomials $\omega_n(x), \ldots, \omega_{n+m}(x)$, and each of these is orthogonal with weight $p(x)$ to lower powers of x.

Equation (110) can be written in the form

$$\int_a^b p(x)\sigma(x)\frac{Q_n(x)}{\sigma(x)}R_i(x)\,dx = 0,$$

whence it follows that the polynomial $\dfrac{Q_n(x)}{\sigma(x)}$ of degree n is orthogonal with weight $p(x)\sigma(x)$ to all polynomials of lower degree. This is possible only if this polynomial differs from $\varphi(x)$ at most by a constant factor; therefore

$$\varphi_n(x) = K\frac{Q_n(x)}{\sigma(x)}. \tag{111}$$

The factor K can be easily found from the normalization condition for $\varphi_n(x)$:

$$K^2\int_a^b p(x)\sigma(x)\left[\frac{Q_n(x)}{\sigma(x)}\right]^2 dx = 1. \tag{112}$$

Remark 1. To complete the proof we have still to convince ourselves that $Q_n(x)$ is not identically zero. Directing our attention to the minor of the element $\omega_{n+m}(x)$, we see that to show this the following lemma is sufficient.

Lemma. *If the real or pairwise complex conjugate numbers $\xi_1, \xi_2, \ldots, \xi_m$ are distinct and lie outside the interval (a, b), then the determinant*

$$\Delta = \begin{vmatrix} \omega_n(\xi_1) & \omega_n(\xi_2) & \cdots & \omega_n(\xi_m) \\ \omega_{n+1}(\xi_1) & \omega_{n+1}(\xi_2) & \cdots & \omega_{n+1}(\xi_m) \\ \cdots & \cdots & \cdots & \cdots \\ \omega_{n+m-1}(\xi_1) & \omega_{n+m-1}(\xi_2) & \cdots & \omega_{n+m-1}(\xi_m) \end{vmatrix}$$

is different from zero.[13]

[13] The following proof was communicated to me by D. K. FADEEV.

Proof. We first show that if a determinant whose columns are real or pairwise complex conjugate [14] is equal to zero, then between its rows there is a linear dependence with real coefficients not all zero. To be specific, let all the columns except the second and third be real:

$$D = \begin{vmatrix} a_1 & b_1 + ic_1 & b_1 - ic_1 & \cdots & r_1 \\ a_2 & b_2 + ic_2 & b_2 - ic_2 & \cdots & r_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n & b_n + ic_n & b_n - ic_n & \cdots & r_n \end{vmatrix} = 0.$$

Adding the third column to the second, factoring 2 from the second, and subtracting finally the second from the third, we obtain the equation

$$\begin{vmatrix} a_1 & b_1 & c_1 & \cdots & r_1 \\ a_2 & b_2 & c_2 & \cdots & r_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_n & b_n & c_n & \cdots & r_n \end{vmatrix} = 0,$$

where this determinant now contains only real elements.

It is therefore possible to find n real numbers $A_1, A_2, \ldots, A_n$ not all zero which satisfy the system of equations

$$\begin{aligned} A_1a_1 + A_2a_2 + \cdots + A_na_n &= 0, \\ A_1b_1 + A_2b_2 + \cdots + A_nb_n &= 0, \\ A_1c_1 + A_2c_2 + \cdots + A_nc_n &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ A_1r_1 + A_2r_2 + \cdots + A_nr_n &= 0. \end{aligned}$$

But then the following equations also hold:

$$\sum A_k a_k = 0, \quad \sum A_k(b_k + ic_k) = 0, \quad \sum A_k(b_k - ic_k) = 0, \ \ldots, \quad \sum A_k r_k = 0.$$

With these preliminaries in mind we now assume $\Delta = 0$. Then it is possible to find m real numbers $A_1, A_2, \ldots, A_m$ not all zero such that

$$A_1\omega_n(\xi_k) + A_2\omega_{n+1}(\xi_k) + \cdots + A_m\omega_{n+m-1}(\xi_k) = 0 \quad (k = 1, 2, \ldots, m).$$

We now define

$$H(x) = A_1\omega_n(x) + A_2\omega_{n+1}(x) + \cdots + A_m\omega_{n+m-1}(x).$$

This is a polynomial of H_{n+m-1} which does not vanish identically, which is

[14] A sequence of numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ is for brevity called real if all the α_k are real. Similarly, two sequences $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$ are called complex conjugate if for all indices k the equation $\alpha_k = \bar{\beta}_k$ holds.

orthogonal to all polynomials of H_{n-1}, and which has all the numbers ξ_1, ξ_2, $\ldots$, ξ_m as roots. It is therefore divisible by

$$\sigma(x) = (x - \xi_1)(x - \xi_2)\cdots(x - \xi_m),$$

and the degree of the quotient is not greater than $n - 1$. If we denote this quotient by $q(x)$, then we obtain

$$\int_a^b p(x)H(x)q(x)\,dx = 0,$$

which is however impossible, since the product

$$H(x)q(x) = \sigma(x)q^2(x)$$

does not change sign on $[a, b]$. The lemma has now been proved.

Remark 2. One could still doubt that it is possible to compute K from equation (112), since $Q_n(x)$ is a polynomial with complex coefficients such that

$$\int_a^b p(x)\sigma(x)\left[\frac{Q_n(x)}{\sigma(x)}\right]^2 dx = 0 \tag{113}$$

might also occur.

The equation (113) is however impossible, since the fraction $\dfrac{Q_n(x)}{\sigma(x)}$ can be represented in the form $\dfrac{\varphi_n(x)}{K}$.

Remark 3. In constructing the polynomials $\varphi_n(x)$ we have used only the orthogonality of the polynomials $\omega_n(x)$ with weight $p(x)$ but have made no use of their normalization.

Example. Let us construct the polynomials $U_n(x)$ which form an orthonormal system with weight $\sqrt{1 - x^2}$ on $[-1, +1]$.

Now

$$\sqrt{1 - x^2} = \frac{1 - x^2}{\sqrt{1 - x^2}}.$$

The orthogonal system with weight $\dfrac{1}{\sqrt{1 - x^2}}$ is known to us: it consists of the TCHEBYSHEFF polynomials $T_n(x)$.

The polynomials $U_n(x)$ sought therefore differ only by a constant factor from the polynomials

$$\frac{1}{1-x^2}\begin{vmatrix} T_n(1) & T_n(-1) & T_n(x) \\ T_{n+1}(1) & T_{n+1}(-1) & T_{n+1}(x) \\ T_{n+2}(1) & T_{n+2}(-1) & T_{n+2}(x) \end{vmatrix}. \tag{114}$$

But now $T_n(x) = \cos(n \arccos x)$; therefore

$$T_n(1) = 1, \quad T_n(-1) = (-1)^n,$$

so that (114) becomes

$$(-1)^{n+1} 2 \frac{T_{n+2}(x) - T_n(x)}{1-x^2}.$$

From this it follows that

$$U_n(x) = K \frac{T_{n+2}(x) - T_n(x)}{1-x^2}. \tag{115}$$

We must now find K from the condition

$$\int_{-1}^{+1} \sqrt{1-x^2}\, U_n^2(x)\, dx = 1.$$

Putting in (115) and substituting $x = \cos\theta$, we obtain

$$K = \frac{1}{\sqrt{2\pi}}.$$

From (115) and $|T_n(x)| \leqq 1$ we obtain the estimate

$$|U_n(x)| \leqq \sqrt{\frac{2}{\pi}} \frac{1}{1-x^2}. \tag{116}$$

This estimate is better than that obtained from Theorem 1, for the normalized TCHEBYSHEFF polynomials are $\sqrt{\frac{2}{\pi}}\, T_n(x)$ (for $n > 0$), and estimate (102) therefore gives in our case only that

$$|U_n(x)| \leqq 3\sqrt{\frac{2}{\pi}} \frac{1}{1-x^2}.$$

The polynomials $U_n(x)$ have many important properties. We return to them in Chapter VI. In particular, we will there be able to give an even better estimate than (116).

CHAPTER V

THE LEGENDRE POLYNOMIALS

§1. The RODRIGUES Formula

Definition. The polynomials $X_n(x)$ which form an orthogonal system on the interval $[-1, +1]$ with weight function $p(x) = 1$ are called LEGENDRE *polynomials.*

Since in this definition we have not taken the normalization of the polynomials into account, they have hereby been defined only up to a constant factor. When this factor is so chosen that they are normalized (and have positive leading coefficients), following V. L. GONCHAROV we shall denote them by $X_n(x)$; if the leading coefficient is made equal to one, we denote them by $\tilde{X}_n(x)$.

These polynomials were introduced by LEGENDRE [1] in the year 1785; in 1814 RODRIGUES [1] gave a simple and very useful formula for them.

To derive this formula we integrate $X_n(x)$ successively n times and denote the polynomial of degree $2n$ thus obtained by $u_n(x)$. We choose the constant of integration such that

$$u_n(-1) = u_n'(-1) = \cdots = u_n^{(n-1)}(-1) = 0. \tag{117}$$

The relations (117) together with the equality $u_n^{(n)}(x) = X_n(x)$ completely determine the function $u_n(x)$ up to a constant factor.

Now let $v(x)$ be a polynomial of degree less than n. Then

$$\int_{-1}^{+1} u_n^{(n)}(x)v(x)\,dx = 0. \tag{118}$$

Integration by parts gives

$$\int_{-1}^{+1} u_n^{(n)} v\,dx$$

$$= [u_n^{(n-1)}v - u_n^{(n-2)}v' + \cdots + (-1)^{n-1}u_n v^{(n-1)}]_{-1}^{+1} + (-1)^n \int_{-1}^{+1} u_n v^{(n)}\,dx.$$

Since $v^{(n)}(x) = 0$, (117) and (118) imply

$$u_n^{(n-1)}(1)v(1) - u_n^{(n-2)}(1)v'(1) + \cdots + (-1)^{n-1}u_n(1)v^{(n-1)}(1) = 0.$$

Since now the numbers $v(1), v'(1), \ldots, v^{(n-1)}(1)$ are entirely arbitrary,[1] this implies

$$v(x) = A_0 + \frac{A_1}{1!}(x-1) + \cdots + \frac{A_{n-1}}{(n-1)!}(x-1)^{n-1}.$$

$$u_n(1) = u_n'(1) = \cdots = u_n^{(n-1)}(1) = 0. \tag{119}$$

Equations (117) and (119) state that each of the two points $x = \pm 1$ is an n-fold root of the polynomial $u_n(x)$. This polynomial is therefore divisible by $(x-1)^n(x+1)^n$, but since it is of degree $2n$ this means

$$u_n(x) = K_n(x^2-1)^n$$

and hence

$$X_n(x) = K_n \frac{d^n(x^2-1)^n}{dx^n}. \tag{120}$$

This is the RODRIGUES *formula.*

Since

$$[(x^2-1)^n]^{(n)} = 2n(2n-1)\cdots(n+1)x^n + \cdots$$

it is clear that

$$\tilde{X}_n(x) = \frac{n!}{(2n)!}\frac{d^n(x^2-1)^n}{dx^n}. \tag{121}$$

In order to find the value K_n which normalizes the polynomial $X_n(x)$, we again apply integration by parts where we set $v(x) = u_n^{(n)}(x)$. According to (117) and (119) all the terms outside the integral sign vanish, so that we obtain

$$\int_{-1}^{+1} [u_n^{(n)}(x)]^2\,dx = (-1)^n \int_{-1}^{+1} u_n(x)u_n^{(2n)}(x)\,dx. \tag{122}$$

Here

$$u_n^{(2n)}(x) = K_n(2n)!.$$

On the other hand, putting

$$I_n = \int_{-1}^{+1} (x^2-1)^n\,dx,$$

[1] For any prescribed numbers $A_0, A_1, \ldots, A_{n-1}$ it is always possible to find a polynomial $v(x)$ of degree less than n for which $v^{(k)}(1) = A^k$ $(k = 0, 1, 2, \ldots, n-1)$, namely the polynomial

we have

$$I_n = \int_{-1}^{+1} x^2(x^2 - 1)^{n-1}\, dx - I_{n-1}.$$

Integrating by parts we find

$$\int_{-1}^{+1} x^2(x^2 - 1)^{n-1}\, dx = \frac{1}{2n}\int_{-1}^{+1} x d(x^2 - 1)^n = -\frac{1}{2n} I_n,$$

and therefore

$$I_n = -\frac{2n}{2n+1} I_{n-1}.$$

Replacing n successively by $n - 1, n - 2, \ldots, 1$ and noting that $I_0 = 2$, we obtain

$$I_n = (-1)^n \frac{(2n)!!}{(2n+1)!!} 2.$$

Equation (122) then becomes

$$\int_{-1}^{+1} X_n^2(x)\, dx = \frac{(2n)!(2n)!!}{(2n+1)!!} 2K_n^2,$$

or equivalently

$$\int_{-1}^{+1} X_n^2(x)\, dx = \frac{[(2n)!!]^2}{2n+1} 2K_n^2.$$

If the integral is to be one, then the constant

$$K_n = \frac{1}{(2n)!!}\sqrt{\frac{2n+1}{2}};$$

hence

$$\hat{X}_n(x) = \sqrt{\frac{2n+1}{2}}\, \frac{1}{(2n)!!}\, \frac{d^n(x^2 - 1)^n}{dx^n}.$$

From the general theory we have

Theorem 1. *All the roots of the polynomial $X_n(x)$ are real, distinct, and lie in the interval $(-1, +1)$.*

We can also derive this directly from the RODRIGUES formula without making use of the general theory. Clearly ± 1 are roots of multiplicity n of $u_n(x) = (x^2 - 1)^n$. The mean value theorem therefore asserts the existence of a root ξ_1 of $u_n'(x)$ in the interval $(-1, +1)$. Now ± 1 are roots of multi-

plicity $n-1$ of $u_n'(x)$, and the mean value theorem therefore asserts that there are two roots η_1 and η_2 of $u_n''(x)$ in the intervals $(-1, \xi_1)$ and $(\xi_1, +1)$. Likewise ± 1 are roots of multiplicity $n-2$ of $u_n''(x)$.

For $k < n$ the derivative $u_n^{(k)}(x)$ has exactly k distinct roots in the interval $(-1, +1)$ and the points ± 1 are roots of multiplicity $n-k$. The derivative $u_n^{(k+1)}$ then has $k+1$ distinct roots in this interval by the mean value theorem. Hence the assertion follows.

The RODRIGUES formula provides an explicit representation for $X_n(x)$, since

$$(x^2-1)^n = \sum_{k=0}^{n} (-1)^k C_n^k x^{2n-2k},$$

and therefore

$$X_n(x) = K_n \sum_{k=0}^{\left[\frac{n}{2}\right]} (-1)^k \frac{(2n-2k)!}{(n-2k)!} C_n^k x^{n-2k}. \tag{123}$$

The formula shows that only summands whose exponents are even if n is even and odd if n is odd occur in $X_n(x)$.

Therefore if we apply the generally valid recursion formula

$$\tilde{X}_{n+2}(x) = (x-\alpha_{n+2})\tilde{X}_{n+1}(x) - \lambda_{n+1}\tilde{X}_n(x), \tag{124}$$

to the LEGENDRE polynomials, then $\alpha_{n+2} = 0$.

As for λ_{n+1}, we find this value as in the general theory by multiplying (124) by $\tilde{X}_n(x)$ and then integrating. We obtain as before

$$\int_{-1}^{+1} x\tilde{X}_{n+1}(x)\tilde{X}_n(x)\,dx = \int_{-1}^{+1} \tilde{X}_{n+1}^2(x)\,dx.$$

Therefore

$$\lambda_{n+1} = \frac{\int_{-1}^{+1} \tilde{X}_{n+1}^2(x)\,dx}{\int_{-1}^{+1} \tilde{X}_n^2(x)\,dx}.$$

However, we have seen that

$$\int_{-1}^{+1} \tilde{X}_n^2(x)\,dx = \frac{[(2n)!!]^2}{2n+1} 2K_n^2$$

with $K_n = \dfrac{n!}{(2n)!}$. From this it follows after simple rearrangements that

$$\lambda_{n+1} = \frac{(n+1)^2}{(2n+1)(2n+3)},$$

and hence we have

$$\tilde{X}_{n+2}(x) = x\tilde{X}_{n+1}(x) - \frac{(n+1)^2}{(2n+1)(2n+3)}\tilde{X}_n(x). \tag{125}$$

From the RODRIGUES formula $\tilde{X}_1(x) = x$ and $\tilde{X}_0(x) = 1$.
From this and (125)

$$\begin{aligned}
&\tilde{X}_0(x) = 1,\\
&\tilde{X}_1(x) = x,\\
&\tilde{X}_2(x) = \tfrac{1}{3}(3x^2 - 1),\\
&\tilde{X}_3(x) = \tfrac{1}{5}(5x^3 - 3x),\\
&\tilde{X}_4(x) = \tfrac{1}{35}(35x^4 - 30x^2 + 3),\\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

The recursion formula (125) also makes possible the construction of the continued fraction (83) corresponding to the LEGENDRE polynomials. We note that $\alpha_1 = 0$ and $\lambda_0 = 2$, since α_1 is the root of the polynomial $X_1(x)$ and $\lambda_0 = \int_{-1}^{+1} dx$. Taking into account the formula

$$\int_{-1}^{+1} \frac{dt}{x-t} = \ln\frac{x+1}{x-1} \qquad (x > 1),$$

we obtain

$$\ln\frac{x+1}{x-1} = \cfrac{2}{x - \cfrac{1/3}{x - \cfrac{4/15}{x - \cfrac{9/35}{x - \cdots}}}}.$$

In many problems the polynomials $X_n(x)$ obtained from (120) by putting $K_n = \frac{1}{(2n)!!}$ are of interest; we denote them by $P_n(x)$, so that

$$P_n(x) = \frac{1}{(2n)!!}\frac{d^n(x^2-1)^n}{dx^n}.$$

Obviously

$$\hat{X}_n(x) = \sqrt{\frac{2n+1}{2}}\,P_n(x); \quad \tilde{X}_n(x) = \frac{n!}{(2n-1)!!}P_n(x). \tag{126}$$

For these polynomials

$$A_n = \int_{-1}^{+1} P_n^2(x)\,dx = \frac{2}{2n+1},$$

and the recursion formula for them is especially simple:

$$(n+2)P_{n+2}(x) = (2n+3)xP_{n+1}(x) - (n+1)P_n(x). \tag{127}$$

In addition to formulas (120) and (123) we can derive still a third explicit expression for $X_n(x)$. From the LEIBNIZ formula

$$(uv)^{(n)} = \sum_{k=0}^{n} C_n^k u^{(n-k)} v^{(k)}$$

we have

$$[(x^2-1)^n]^{(n)} = \sum_{k=0}^{n} C_n^k [(x-1)^n]^{(n-k)} [(x+1)^n]^{(k)}$$

$$= \sum_{k=0}^{n} C_n^k \frac{n!}{k!}(x-1)^k \frac{n!}{(n-k)!}(x+1)^{n-k}.$$

From this it follows that

$$X_n(x) = K_n n! \sum_{k=0}^{n} [C_n^k]^2 (x-1)^k (x+1)^{n-k}$$

and in particular

$$P_n(1) = 1, \quad P_n(-1) = (-1)^n. \tag{128}$$

Another consequence of the RODRIGUES formula is

Theorem 2. *The* LEGENDRE *polynomial* $y = X_n(x)$ *satisfies the differential equation*

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0. \tag{129}$$

Proof. Differentiate the equation

$$u = (x^2-1)^n$$

and multiply the result by $x^2 - 1$. This gives

$$u'(x^2-1) = 2nxu.$$

Now differentiate both sides of this equation $n+1$ times. The LEIBNIZ formula gives

$$\sum_{k=0}^{n+1} C_{n+1}^k u^{(n+2-k)} (x^2-1)^{(k)} = 2n \sum_{k=0}^{n+1} C_{n+1}^k u^{(n+1-k)} x^{(k)}$$

and hence

$$u^{(n+2)}(x^2-1) + (n+1)u^{(n+1)}2x + \frac{(n+1)n}{2} 2u^{(n)}$$

$$= 2n[u^{(n+1)}x + (n+1)u^{(n)}].$$

Putting $u^{(n)} = y$ we obtain (129).

In conclusion we note still another identity which relates three successive LEGENDRE polynomials:

$$P'_{n+1}(x) - P'_{n-1}(x) = (2n + 1)P_n(x). \tag{130}$$

To prove this we write the left side in the form

$$\alpha = \left\{\left[\frac{(x^2 - 1)^{n+1}}{(2n + 2)!!}\right]^{(n+1)} - \left[\frac{(x^2 - 1)^{n-1}}{(2n - 2)!!}\right]^{(n-1)}\right\}';$$

$$\alpha = \left\{\left[\frac{(x^2 - 1)^{n+1}}{(2n + 2)!!}\right]'' - \frac{(x^2 - 1)^{n-1}}{(2n - 2)!!}\right\}^{(n)}$$

But now

$$\left[\frac{(x^2 - 1)^{n+1}}{(2n + 2)!!}\right]'' = \frac{(2n + 1)x^2 - 1}{(2n)!!}(x^2 - 1)^{n-1}.$$

From this it follows that

$$\alpha = \left\{\left[\frac{(2n + 1)x^2 - 1}{(2n)!!} - \frac{1}{(2n - 2)!!}\right](x^2 - 1)^{n-1}\right\}^{(n)}$$

$$= \frac{2n + 1}{(2n)!!}[(x^2 - 1)^n]^{(n)},$$

which establishes (130).

§2. The Generating Function

We consider the function

$$X(t, x) = \frac{1}{\sqrt{1 - 2tx + t^2}},$$

which for fixed x and sufficiently small $|t|$ can be developed in a power series. Now in the binomial series

$$(1 - z)^{-\frac{1}{2}} = 1 + \frac{1}{2}z + \frac{3!!}{2^2 2!}z^2 + \frac{5!!}{2^3 3!}z^3 + \cdots \tag{131}$$

all the coefficients are positive. If we here replace z by $2tx - t^2$ we obtain $X(t, x)$ in the form

$$X(t, x) = \sum_{k=0}^{\infty} \alpha_k (2tx - t^2)^k, \tag{132}$$

where all the coefficients α_k are positive. If we set $z = 2|tx| + t^2$ in (131), then the series

$$\sum_{k=0}^{\infty} \alpha_k (2|tx| + t^2)^k \tag{133}$$

converges for $2|tx| + t^2 < 1$. The series (132) can be written as a double series:

$$\begin{aligned} &\alpha_0 + \\ &\quad + \alpha_1 2tx - \quad \alpha_1 t^2 + \\ &\qquad + 4\alpha_2 x^2 t^2 - 4\alpha_2 x t^3 + \alpha_2 t^4 + \\ &\qquad\qquad + \cdots \end{aligned}$$

This double series converges *absolutely* under the above condition, since it is majorized by the positive double series formed similarly from (133). We may therefore rearrange terms, thus obtaining a series of the form

$$X(t, x) = \sum_{n=0}^{\infty} \alpha_n(x) t^n. \tag{134}$$

For $t = 0$ we obtain $\alpha_0(x) = 1$. Differentiating (134) and again putting $t = 0$ we find $\alpha_1(x) = x$. Hence

$$\alpha_0(x) = P_0(x), \quad \alpha_1(x) = P_1(x).$$

We now show that in general $\alpha_n(x) = P_n(x)$. To this end we differentiate (134) with respect to t:

$$\frac{x - t}{(1 - 2tx + t^2)^{\frac{1}{2}}} = \sum_{n=1}^{\infty} n\alpha_n(x) t^{n-1} = \sum_{n=0}^{\infty} (n + 1)\alpha_{n+1}(x) t^n.$$

We multiply this equation by $1 - 2tx + t^2$ and replace $(1 - 2tx + t^2)^{-\frac{1}{2}}$ on the left-hand side by formula (134):

$$(x - t) \sum_{n=0}^{\infty} \alpha_n(x) t^n = (1 - 2tx + t^2) \sum_{n=0}^{\infty} (n + 1)\alpha_{n+1}(x) t^n.$$

Comparing coefficients of t for $n \geqq 1$ we obtain

$$x\alpha_n(x) - \alpha_{n-1}(x) = (n + 1)\alpha_{n+1}(x) - 2nx\alpha_n(x) + (n - 1)\alpha_{n-1}(x)$$

and hence

$$(n + 1)\alpha_{n+1}(x) = (2n + 1)x\alpha_n(x) - n\alpha_{n-1}(x) \qquad (n \geqq 1).$$

Replacing n by $n + 1$ we arrive at the recursion formula

$$(n + 2)\alpha_{n+2}(x) = (2n + 3)x\alpha_{n+1}(x) - (n + 1)\alpha_n(x) \qquad (n \geqq 0),$$

which is exactly of the form of recursion formula (127) for the polynomials $P_n(x)$. We have thus proved

Theorem 1.

$$(1 - 2tx + t^2)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} P_n(x) t^n. \tag{135}$$

The function on the left-hand side is called the *generating function* of the LEGENDRE polynomials.

Using formula (135) an estimate of the polynomials $P_n(x)$ can be obtained:

Theorem 2. *The inequality*

$$|P_n(x)| \leqq 1 \tag{136}$$

holds on the interval $[-1, +1]$.

Proof. If $x \in [-1, +1]$ we set $x = \cos\theta$. From formula (135)

$$(1 - 2t\cos\theta + t^2)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} P_n(\cos\theta)t^n.$$

Now

$$1 - 2t\cos\theta + t^2 = (1 - te^{i\theta})(1 - te^{-i\theta}).$$

From the binomial formula (131) it follows that

$$(1 - 2t\cos\theta + t^2)^{-\frac{1}{2}} = \left[1 + \sum_{k=1}^{\infty} \frac{(2k-1)!!}{(2k)!!} t^k e^{ik\theta}\right]\left[1 + \sum_{k=1}^{\infty} \frac{(2k-1)!!}{(2k)!!} t^k e^{-ik\theta}\right].$$

Multiplying these series in the usual manner we obtain for the coefficients $P_n(\cos\theta)$ of t^n

$$P_n(\cos\theta) = \frac{(2n-1)!!}{(2n)!!} e^{in\theta} + \sum_{k=1}^{n-1} \frac{(2k-1)!!}{(2k)!!} \frac{[2(n-k)-1]!!}{[2(n-k)]!!} e^{i(2k-n)\theta} + \frac{(2n-1)!!}{(2n)!!} e^{-in\theta}.$$

It is important that all the coefficients on the right-hand side are positive; hence

$$|P_n(\cos\theta)| \leqq \frac{(2n-1)!!}{(2n)!!} + \sum_{k=1}^{n-1} \frac{(2k-1)!!}{(2k)!!} \frac{(2n-2k-1)!!}{(2n-2k)!!} + \frac{(2n-1)!!}{(2n)!!}.$$

But now the right-hand side of this inequality is just $P_n(\cos 0) = P_n(1) = 1$. The theorem is herewith proved. We shall give still other estimates for the LEGENDRE polynomials in the next section.

As a further example of the usefulness of formula (135), we shall use it to derive three other relations among the LEGENDRE polynomials. Differentiation of formula (135) once with respect to x and once with respect to t gives

$$t(1 - 2tx + t^2)^{-3/2} = \sum_{n=1}^{\infty} P_n'(x)t^n,$$

$$(x - t)(1 - 2tx + t^2)^{-3/2} = \sum_{n=1}^{\infty} P_n(x)nt^{n-1},$$

whence

$$(x - t)\sum_{n=1}^{\infty} P_n'(x)t^n = \sum_{n=1}^{\infty} P_n(x)nt^n$$

or equivalently [2]

$$\sum_{n=1}^{\infty} xP_n'(x)t^n - \sum_{n=1}^{\infty} P_{n-1}'(x)t^n = \sum_{n=1}^{\infty} nP_n(x)t^n.$$

Comparing coefficients gives the first of the relations in which we are interested:

$$xP_n'(x) - P_{n-1}'(x) = nP_n(x). \tag{137}$$

Subtraction of this equation from (130) gives

$$P_{n+1}'(x) - xP_n'(x) = (n + 1)P_n(x).$$

Finally we here replace n by $n - 1$ and eliminate $P_{n-1}'(x)$ from the equation so obtained by means of (137). We thus find the third of the relations sought:

$$(1 - x^2)P_n'(x) = nP_{n-1}(x) - nxP_n(x).$$

§3. The LAPLACE Integral

We consider the function

$$y_n(x) = \frac{1}{\pi}\int_0^{\pi} [x + i\sqrt{1 - x^2}\cos\theta]^n d\theta.$$

Since $y_0(x) = 1$ and $y_1(x) = x$, it follows that

$$y_0(x) = P_0(x), \quad y_1(x) = P_1(x).$$

We now show that in general $y_n(x) = P_n(x)$. For this we clearly have only to show that the functions $y_n(x)$ satisfy the same recursion formula as the polynomials $P_n(x)$, i.e.

$$(n + 2)y_{n+2} - (2n + 3)xy_{n+1} + (n + 1)y_n = 0. \tag{138}$$

For this purpose we temporarily put $\alpha = x + i\sqrt{1 - x^2}\cos\theta$. Then

$$y_n(x) = \frac{1}{\pi}\int_0^{\pi} \alpha^n \, d\theta,$$

$$y_{n+1}(x) = \frac{1}{\pi}\int_0^{\pi} [x + i\sqrt{1-x^2}\cos\theta]\alpha^n \, d\theta,$$

[2] In the second summation we may sum from $n = 1$ on (instead of from $n = 2$), since $P_{n-1}'(x) = 0$ for $n = 1$.

$$y_{n+2}(x) = \frac{1}{\pi}\int_0^{\pi} [x + i\sqrt{1-x^2}\cos\theta]^2\alpha^n\, d\theta.$$

The left side of (138) now assumes the form

$$\frac{1}{\pi}\int_0^{\pi} \beta\alpha^n\, d\theta,$$

where

$$\beta = (n+2)(x + i\sqrt{1-x^2}\cos\theta)^2 - (2n+3)x(x + i\sqrt{1-x^2}\cos\theta) + n + 1.$$

Simple rearrangements yield

$$\beta = (n+1)(1-x^2)\sin^2\theta + i\sqrt{1-x^2}\,(x + i\sqrt{1-x^2}\cos\theta)\cos\theta.$$

If we denote

$$(n+1)(1-x^2)\sin^2\theta \text{ by } \gamma,$$

$$i\sqrt{1-x^2}\,(x + i\sqrt{1-x^2}\cos\theta)\cos\theta \text{ by } \delta,$$

then $\beta = \gamma + \delta$. But now

$$\int_0^{\pi} \delta\alpha^n\, d\theta = i\sqrt{1-x^2}\int_0^{\pi} \alpha^{n+1}\cos\theta\, d\theta.$$

Integration by parts gives

$$\int_0^{\pi} \delta\alpha^n\, d\theta = i\sqrt{1-x^2}\left\{[\alpha^{n+1}\sin\theta]_0^{\pi} - (n+1)\int_0^{\pi} \alpha^n\,\alpha'\sin\theta\, d\theta\right\}.$$

Since $\alpha' = -i\sqrt{1-x^2}\sin\theta$, it follows from this that

$$\int_0^{\pi} \delta\alpha^n\, d\theta = -(n+1)(1-x^2)\int_0^{\pi} \alpha^n\sin^2\theta\, d\theta = -\int_0^{\pi} \gamma\alpha^n\, d\theta$$

and hence

$$\int_0^{\pi} \beta\alpha^n\, d\theta = 0,$$

which proves our assertion. The LEGENDRE polynomials therefore have the integral representation

$$P_n(x) = \frac{1}{\pi}\int_0^{\pi} [x + i\sqrt{1-x^2}\cos\theta]^n\, d\theta; \tag{139}$$

The integral itself is called a LAPLACE *integral*. For $-1 \leqq x \leqq 1$

$$|x + i\sqrt{1-x^2}\,\theta| = \sqrt{x^2 + (1-x^2)\cos^2\theta} \leqq 1;$$

and for these x therefore

$$|P_n(x)| \leqq \frac{1}{\pi}\int_0^{\pi} |x + i\sqrt{1-x^2}\cos\theta|^n \, d\theta \leqq 1,$$

and we have again obtained the estimate (136). For the *interior* points of the interval $(-1, +1)$ we can find a closer estimate. Writing $|x + i\sqrt{1-y^2}\cos\theta|$ in the form

$$\sqrt{1-(1-x^2)\sin^2\theta},$$

we find

$$|P_n(x)| \leqq \frac{1}{\pi}\int_0^{\pi} [1-(1-x^2)\sin^2\theta]^{\frac{n}{2}} \, d\theta.$$

We decompose this integral into two parts extending over the intervals $\left[0, \frac{\pi}{2}\right]$ and $\left[\frac{\pi}{2}, \pi\right]$, whereby in so doing we replace θ by $\pi - \theta$ in the second integral. This gives the inequality

$$|P_n(x)| \leqq \frac{2}{\pi}\int_0^{\pi/2} [1-(1-x^2)\sin^2\theta]^{\frac{n}{2}} \, d\theta.$$

Now as we know, for $0 \leqq \theta \leqq \frac{\pi}{2}$

$$\sin\theta \geqq \frac{2}{\pi}\theta;$$

hence

$$1-(1-x^2)\sin^2\theta \leqq 1 - \frac{4}{\pi_2}(1-x^2)\theta^2.$$

Now for $\alpha > 0$

$$1 - \alpha < e^{-\alpha}.$$

(For if $\alpha > 0$ the derivative of the function $\varphi(\alpha) = e^{-\alpha} + \alpha - 1$ is positive and therefore $\varphi(\alpha) > \varphi(0) = 0$).

From this it follows that

$$|P_n(x)| \leqq \frac{2}{\pi}\int_0^{\pi/2} e^{-\frac{2n}{\pi^2}(1-x^2)\theta^2} \, d\theta,$$

and hence a fortiori

$$|P_n(x)| \leqq \frac{2}{\pi}\int_0^{+\infty} e^{-\frac{2n}{\pi^2}(1-x)^2\theta^2} \, d\theta.$$

Setting $\theta = \dfrac{\pi}{\sqrt{2n(1-x^2)}}\, t$ we obtain

$$|P_n(x)| \leqq \sqrt{\frac{2}{n}}\,\frac{1}{\sqrt{1-x^2}}\int_0^{+\infty} e^{-t^2}\,dt,$$

and since

$$\int_0^{+\infty} e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}$$

we obtain finally

$$|P_n(x)| \leqq \sqrt{\frac{\pi}{2n}}\,\frac{1}{\sqrt{1-x^2}} \quad (-1 < x < 1). \tag{140}$$

§4. Expansions in Terms of LEGENDRE Polynomials

In view of the relations

$$|P_n(x)| \leqq 1, \qquad \hat{X}_n(x) = \sqrt{\frac{2n+1}{2}}\,P_n(x)$$

the kernel $K_n(t, x)$ of the system of LEGENDRE polynomials is subject to the estimate

$$|K_n(t, x)| = \left|\sum_{k=0}^{n} \hat{X}_k(t)\hat{X}_k(x)\right| \leqq \sum_{k=0}^{n}\frac{2k+1}{2} = \frac{(n+1)^2}{2}.$$

For the corresponding LEBESGUE function

$$L_n(x) = \int_{-1}^{+1} |K_n(t, x)|\,dt$$

it thus follows that

$$L_n(x) \leqq (n+1)^2.$$

From Theorem 7 of § 4, Chapter IV it follows that every continuous function $f(x)$ whose best approximation satisfies the condition

$$\lim_{n\to\infty} n^2 E_n(f) = 0 \tag{141}$$

can be expanded in a FOURIER series of LEGENDRE polynomials, and the series is moreover uniformly convergent on the entire interval $[-1, +1]$. From the JACKSON theorem (Vol. 1, Ch. VI, § 2) condition (141) is fulfilled whenever the function possesses a continuous second derivative on the interval. Hence we have

Theorem 1. *If a function $f(x)$ defined on the interval $[-1, +1]$ has there a continuous second derivative $f''(x)$, then it can be expanded in a* FOURIER *series of* LEGENDRE *polynomials which converges uniformly on the entire interval.*

If we do not require that $f(x)$ be expandable on the entire interval $[-1, +1]$, but rather restrict ourselves to the interval $[-1 + h, 1 - h]$ where $0 < h < 1$, then we can considerably relax the conditions on $f(x)$. This is because on the interval $[-1 + h, 1 - h]$ we may replace the estimate $|P_n(x)| \leqq 1$ (which cannot be improved on the interval $[-1, +1]$, since $P_n(x) = 1$) by

$$|P_n(x)| \leqq \sqrt{\frac{\pi}{2n}}\,\frac{1}{h}, \tag{142}$$

which is obtained from (140).

Theorem 2. *If a function satisfies the* DINI-LIPSCHITZ *condition*

$$\lim_{\delta \to 0} \omega(\delta) \ln \delta = 0, \tag{143}$$

on the entire interval $[-1, +1]$, then it can be expanded in a FOURIER *series of* LEGENDRE *polynomials at all points of the interval $(-1, +1)$, and this series converges uniformly in each interval $[-1 + h, 1 - h]$.*

Proof. We decompose the integral

$$L_n(x) = \int_{-1}^{+1} |K_n(t, x)|\, dt$$

into five integrals according to the scheme:

$$L_n(x) = \int_{-1}^{-1+\frac{h}{2}} + \int_{-1+\frac{h}{2}}^{x-\frac{1}{n}} + \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} + \int_{x+\frac{1}{n}}^{1-\frac{h}{2}} + \int_{1-\frac{h}{2}}^{1} = I_1 + I_2 + I_3 + I_4 + I_5.$$

We now estimate integrals I_1 and I_5. From the CHRISTOFFEL-DARBOUX formula

$$K_n(t, x) = \theta_n \frac{\hat{X}_{n+1}(t)\hat{X}_n(x) - \hat{X}_{n+1}(x)\hat{X}_n(t)}{t - x} \quad (0 \leqq \theta_n \leqq 1).$$

Now if $-1 + h \leqq x \leqq 1 - h$ and $-1 \leqq t \leqq -1 + \frac{h}{2}$, then $|t - x| \geqq \frac{h}{2}$.

Moreover, according to (142)

$$|\hat{X}_n(x)| \leqq \sqrt{\frac{2n+1}{2}}\,\sqrt{\frac{\pi}{2n}}\,\frac{1}{h} < \frac{\sqrt{\pi}}{h}$$

is an estimate which is also valid for $|\hat{X}_{n+1}(x)|$. Hence

$$I_1 \leqq \frac{2\sqrt{\pi}}{h^2}\left[\int_{-1}^{+1}|\hat{X}_n(t)|\,dt + \int_{-1}^{+1}|\hat{X}_{n+1}(t)|\,dt\right].$$

From the SCHWARZ inequality

$$\int_{-1}^{+1}|\hat{X}_n(t)|\,dt \leqq \sqrt{2},$$

and therefore

$$I_1 \leqq \frac{4\sqrt{2\pi}}{h^2};$$

this estimate also applies to I_5.

In estimating I_2 and I_4 it suffices to consider one of the two integrals; let us take I_4. Here

$$|\hat{X}_n(x)| < \frac{\sqrt{\pi}}{h}, \quad |\hat{X}_n(t)| < \frac{2\sqrt{\pi}}{h},$$

and therefore

$$|K_n(t, x)| \leqq \frac{4\pi}{h^2(t-x)},$$

whence it follows that

$$I_4 \leqq \frac{4\pi}{h^2}\int_{x+\frac{1}{n}}^{1-\frac{h}{2}}\frac{dt}{t-x} = \frac{4\pi}{h^2}\left[\ln n + \ln\left(1 - \frac{h}{2} - x\right)\right].$$

Since

$$\ln\left(1 - \frac{h}{2} - x\right) \leqq \ln 2,$$

we obtain

$$I_4 < \frac{4\pi}{h^2}\ln n + \frac{4\pi}{h^2}\ln 2,$$

which is an estimate that is also valid for I_2.

Finally we estimate I_3. We have

$$|\hat{X}_k(x)| < \frac{\sqrt{\pi}}{h}, \quad |\hat{X}_k(t)| < \frac{2\sqrt{\pi}}{h};$$

therefore

$$|K_n(t, x)| \leqq \sum_{k=0}^{n}|\hat{X}_k(t)\hat{X}_k(x)| \leqq \frac{2\pi}{h^2}(n+1)$$

and hence

$$I_3 = \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} |K_n(t, x)|\, dt \leqq \frac{4\pi(n+1)}{nh^2} < \frac{8\pi}{h^2}.$$

All of these together give an estimate for $L_n(x)$

$$L_n(x) < \frac{8\sqrt{2\pi}}{h^2} + \frac{8\pi}{h^2} \ln n + \frac{8\pi}{h^2} + \frac{8\pi}{h^2} \ln 2. \tag{144}$$

If condition (143) is fulfilled, then from the JACKSON theorem

$$\lim_{n\to\infty} E_n(f) \ln n = 0,$$

and hence according to (144) also

$$\lim_{n\to\infty} E_n(f) L_n(x) = 0,$$

wherewith the theorem is proved.

Theorems 1 and 2 require of the function to be expanded certain properties of its differentiability on the interval $[-1, +1]$ from which the expandability of the function on the entire interval $[-1, +1]$ or at least on the interval $(-1, +1)$ then follows; neither of the theorems are of local character. Now since at each point x of the interval $(-1, +1)$ the polynomials $\hat{X}_n(x)$ are uniformly bounded (with respect to n), i.e.

$$|\hat{X}_n(x)| \leqq \frac{\sqrt{\pi}}{2} \sqrt{\frac{2n+1}{n}} \frac{1}{\sqrt{1-x^2}} \leqq \frac{\sqrt{3\pi}}{2\sqrt{1-x^2}} \tag{145}$$

starting with Theorem 2 of § 4, Chapter IV we can obtain the following purely local theorem:

Theorem 3. *A function $f(x)$ of the class L^2 can be expanded in a* FOURIER *series of* LEGENDRE *polynomials at each point x of the interval $(-1, +1)$ where the condition*

$$\int_{-1}^{+1} \left[\frac{f(t) - f(x)}{t - x}\right]^2 dt < +\infty \tag{146}$$

is satisfied.

In particular, this condition is fulfilled whenever a finite derivative $f'(x)$ exists at x.

Estimate (145) makes it possible to carry over the general localization principle (Theorem 5, § 4, Chapter IV) to the LEGENDRE polynomials:

Theorem 4. *If two functions $f(x)$ and $g(x)$ of L^2 coincide in the interval $(x_0 - h, x_0 + h)$, then*

$$\lim_{n\to\infty} \{S_n[f, x_0] - S_n[g, x_0]\} = 0.$$

Here $S_n[f, x]$ and $S_n[g, x]$ are the partial sums of the expansions in terms of LEGENDRE polynomials.

Theorem 5. *Let $f(t)$ be a function of the class L^2 such that at each point x of the interval $(-1, +1)$ the limits $f(x - 0)$ and $f(x + 0)$ exist. If both of the integrals*

$$\int_{-1}^{x} \left[\frac{f(t) - f(x - 0)}{t - x}\right]^2 dt, \quad \int_{x}^{+1} \left[\frac{f(t) - f(x + 0)}{t - x}\right]^2 dt \tag{147}$$

are finite, then the expansion of the function $f(x)$ in a FOURIER *series of* LEGENDRE *polynomials converges to the sum*

$$\frac{f(x - 0) + f(x + 0)}{2}.$$

Proof. Let us first consider the very simple function $h(t)$ which is equal to 1 for $-1 \leqq t \leqq x$ and equal to 0 for $x < t \leqq 1$. Its FOURIER coefficients are

$$c_k = \int_{-1}^{x} \hat{X}_k(t)\, dt.$$

According to (130), for $k \geqq 1$

$$\hat{X}_k(t) = \sqrt{\frac{2k + 1}{2}}\, P_k(t) = \frac{1}{\sqrt{4k + 2}} [P'_{k+1}(t) - P'_{k-1}(t)],$$

whence

$$c_k = \frac{1}{\sqrt{4k + 2}} [P_{k+1}(x) - P_{k-1}(x)],$$

since $P_{k+1}(-1) = P_{k-1}(-1) = (-1)^{k-1}$.

Now $\hat{X}_0(x) = \frac{1}{\sqrt{2}}$ and $c_0 = \frac{x + 1}{\sqrt{2}}$. The partial sum of the FOURIER series of $h(t)$ at the point x is therefore

$$S_n(x) = \frac{1 + x}{2} + \sum_{k=1}^{n} \frac{1}{\sqrt{4k + 2}} [P_{k+1}(x) - P_{k-1}(x)] \hat{X}_k(x).$$

Since $\hat{X}_k(x) = \sqrt{\frac{2k+1}{2}}\, P_k(x)$ it follows that

$$S_n(x) = \frac{1+x}{2} + \frac{1}{2}\sum_{k=1}^{n} [P_{k+1}(x)P_k(x) - P_k(x)P_{k-1}(x)]$$

$$= \frac{1}{2}[1 + x + P_{n+1}(x)P_n(x) - P_1(x)P_0(x)].$$

Recalling that $P_0(x) = 1$ and $P_1(x) = x$, we obtain finally

$$S_n(x) = \frac{1}{2} + \frac{1}{2}P_{n+1}(x)P_n(x)$$

and hence

$$\lim_{n\to\infty} S_n(x) = \frac{1}{2}.$$

On the other hand, the sum $S_n(x)$ can also be computed from the formula

$$S_n(x) = \int_{-1}^{x} K_n(t, x)\, dt.$$

Hence

$$\lim_{n\to\infty} \int_{-1}^{x} K_n(t, x)\, dt = \frac{1}{2}, \tag{148}$$

and since the integral of the kernel $K_n(t, x)$ over the entire interval $[-1, +1]$ is equal to one, we also have

$$\lim_{n\to\infty} \int_{x}^{1} K_n(t, x)\, dt = \frac{1}{2}.$$

We now proceed to the function $f(t)$ and write its FOURIER partial sum in the form

$$\int_{-1}^{x} f(t)K_n(t, x)\, dt + \int_{x}^{1} f(t)K_n(t, x)\, dt.$$

To prove the theorem it suffices to show that

$$\lim_{n\to\infty} \int_{-1}^{x} f(t)K_n(t, x)\, dt = \frac{1}{2}f(x - 0)$$

and

$$\lim_{n\to\infty} \int_{x}^{1} f(t)K_n(t, x)\, dt = \frac{1}{2}f(x + 0).$$

It is sufficient to prove the first equality; the proof for the second is entirely analogous. On the basis of (148) this amounts to proving

$$\lim_{n\to\infty} \int_{-1}^{x} [f(t) - f(x-0)]K_n(t, x)\, dt = 0. \tag{149}$$

From the CHRISTOFFEL-DARBOUX formula

$$K_n(t, x) = \theta_n \frac{\hat{X}_{n+1}(t)\hat{X}_n(x) - \hat{X}_{n+1}(x)\hat{X}_n(t)}{t - x} \quad (0 \leqq \theta_n \leqq 1).$$

The integral in (149) can therefore be written in the form

$$\theta_n[d_{n+1}\hat{X}_n(x) - d_n\hat{X}_{n+1}(x)],$$

where

$$d_n = \int_{-1}^{x} \frac{f(t) - f(x-0)}{t - x} \hat{X}_n(t)\, dt,$$

and d_{n+1} is similarly defined. The numbers d_n are the FOURIER coefficients of a function which is equal to $\frac{f(t) - f(x-0)}{t - x}$ for $-1 \leqq t \leqq x$ and equal to zero for $x < t \leqq 1$. This function belongs to L^2 by hypothesis, so that its FOURIER coefficients tend to zero with increasing n. Moreover, the $\hat{X}_n(x)$ are bounded, whence we obtain (149) and therewith the proof of the theorem.

The estimates

$$|\hat{X}_n(x)| \leqq \sqrt{\frac{2n+1}{2}} \qquad (-1 \leqq x \leqq 1),$$

$$|\hat{X}_n(x)| < \frac{c}{\sqrt{1 - x^2}} \qquad (-1 < x < 1)$$

for the LEGENDRE polynomials give us several more results which concern expansion with respect to polynomials $\varphi_n(x)$ of an orthogonal system with a weight $\sigma(x)$ different from one. If in particular $\sigma(x)$ is a function satisfying a LIPSHITZ condition and having a positive lower bound on $[-1, +1]$ (e.g., $\sigma(x) = Q(x)$ or $\sigma(x) = \frac{1}{Q(x)}$ where $Q(x)$ is a polynomial with no roots in $[-1, +1]$), then according to the theorems of § 5, Chapter IV the estimates

$$|\varphi_n(x)| \leqq A\sqrt{n} \qquad (-1 \leqq x \leqq 1), \tag{150}$$

$$|\varphi_n(x)| \leqq \frac{A}{\sqrt{1 - x^2}} \qquad (-1 < x < 1),$$

where A is a constant hold for the polynomials $\varphi_n(x)$. From this follows

Theorem 6. *If a function $f(x)$ has a continuous second derivative on the interval $[-1, +1]$, then it can be expanded into a* FOURIER *series in terms of the polynomials $\varphi_n(x)$ which converges uniformly on this interval. If on the other hand $f(x)$ has only a first derivative and satisfies a* LIPSHITZ *condition with exponent $\alpha > \frac{1}{2}$, then it can be expanded in a* FOURIER *series in terms of the polynomials $\varphi_n(x)$ on the interval $(-1, +1)$ which converges uniformly on every interval $[-1 + h, 1 - h]$.*

Proof. From what has been said the estimates

$$|K_n(t, x)| \leqq A_1 n^2 \qquad (-1 \leqq x \leqq 1, \quad -1 \leqq t \leqq 1),$$

$$|K_n(t, x)| \leqq \frac{A_1}{\sqrt{1 - x^2}} \sum_{k=1}^{n} \sqrt{k} \qquad (-1 < x < 1, \quad -1 \leqq t \leqq 1)$$

hold for the kernel $K_n(t, x)$ with $n \geqq 1$, where A_1 is a new constant.[3] For the LEBESGUE function

$$L_n(x) = \int_{-1}^{+1} |K_n(t, x)| \, dt$$

therefore:

$$L_n(x) \leqq 2A_1 n^2 \qquad (-1 \leqq x \leqq 1)$$

$$L_n(x) \leqq \frac{2A_1}{\sqrt{1 - x^2}} n^{3/2} \qquad (-1 < x < 1)$$

From this with the help of the results of § 4, Chapter IV and the JACKSON theorem our assertions follow.

[3] The kernel contains the summands $\varphi_0(t)\varphi_0(x)$ to which inequality (150) cannot be applied. This requires that the constants be made larger.

CHAPTER VI

THE JACOBI POLYNOMIALS

§ 1. Generalization of the RODRIGUES Formula

Denfiition. Those polynomials which form an orthogonal system on the interval $[-1, +1]$ with weight function

$$p(x) = (1 - x)^{\alpha}(1 + x)^{\beta}$$

are called JACOBI *polynomials*. Since we always require that the weight be summable, the exponents α and β must satisfy the conditions

$$\alpha > -1, \quad \beta > -1.$$

Just as for the LEGENDRE polynomials, the JACOBI polynomials are determined only up to a constant factor. When this factor is not specified we shall denote the JACOBI polynomials by $J_n^{(\alpha,\beta)}(x)$. When the factor is so chosen that the leading coefficient of the polynomials is equal to one we shall use the notation $\tilde{J}_n^{(\alpha,\beta)}(x)$, and the normalized JACOBI polynomials with positive leading coefficient will be denoted by $\dot{J}_n^{(\alpha,\beta)}(x)$. We will subsequently make use of another special choice of the factor, but this is deferred until later.

One sees immediately that the LEGENDRE polynomials are a particular class of JACOBI polynomials obtained for $\alpha = \beta = 0$; the TSHEBYSHEFF polynomials are obtained for $\alpha = \beta = -\frac{1}{2}$. We will later consider the so-called TSHEBYSHEFF polynomials of second kind which are obtained for $\alpha = \beta = +\frac{1}{2}$. Quite in general, the case of equal exponents $\alpha = \beta$ exhibits certain special features; the corresponding JACOBI polynomials are called *ultraspherical* polynomials.

The formula

$$J_n^{(\alpha,\beta)}(x) = K_n(1 - x)^{-\alpha}(1 + x)^{-\beta}\frac{d^n}{dx^n}[(1 - x)^{\alpha+n}(1 + x)^{\beta+n}] \quad (151)$$

is of great use in the study of the JACOBI polynomials. This formula is a direct generalization of the RODRIGUES formula into which it goes over for $\alpha = \beta = 0$.

To prove this formula we use the LEIBNIZ formula to form the expression

$$\frac{d^n}{dx^n}[(1 - x)^{\alpha+n}(1 + x)^{\beta+n}] = \sum_{k=0}^{n} h_k(1 - x)^{\alpha+n-k}(1 + x)^{\beta+k}],$$

where the coefficients h_k are certain constants. Hence

$$\frac{d^n}{dx^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}] = (1-x)^{\alpha}(1+x)^{\beta}y_n(x), \tag{152}$$

where $y_n(x)$ is a polynomial of degree at most n. We will soon show that the degree of $y_n(x)$ is exactly n, but for the time being we note simply that $y_n(x)$ is not identically zero.[1]

From the LEIBNIZ formula again with $m < n$

$$\varphi_m(x) = \frac{d^m}{dx^m}[(1-x)^{\alpha+n}(1+x)^{\beta+n}] = \sum_{k=0}^{m} h'_k(1-x)^{\alpha+n-k}(1+x)^{\beta+n-m+k}.$$

In the last sum all the exponents $\alpha + n - k$ and $\beta + n - m + k$ are positive; hence

$$\varphi_m(-1) = \varphi_m(+1) = 0. \tag{153}$$

We now consider the integral

$$J = \int_{-1}^{+1}(1-x)^{\alpha}(1+x)^{\beta}v(x)y_n(x)\,dx, \tag{154}$$

where $v(x)$ is an arbitrary polynomial of degree less than n. Using (152) we can write this in the form

$$J = \int_{-1}^{+1}u^{(n)}(x)v(x)\,dx,$$

where we have put

$$u(x) = (1-x)^{\alpha+n}(1+x^{\beta+n}).$$

Integration by parts now gives

$$J = [u^{(n-1)}v - u^{(n-2)}v' + \cdots + (-1)^{(n-1)}uv^{(n-1)}]_{-1}^{+1} + (-1)^n\int_{-1}^{+1}uv^{(n)}\,dx. \tag{155}$$

According to (153) the sum preceding the integral on the right-hand side is equal to zero, and since in addition $v^{(n)}(x) = 0$, it follows that

$$J = 0.$$

Therefore $y_n(x)$ is orthogonal with weight function

$$(1-x)^{\alpha}(1+x)^{\beta}$$

[1] If $y_n(x) \equiv 0$, then $(1-x)^{\alpha+n}(1+x)^{\beta+n}$ would be a polynomial from H_{n-1}. But this is impossible, since the only roots of the function $(1-x)^{\alpha+n}(1+x)^{\beta+n}$ are the two points ± 1. If this function were a polynomial from H_{n-1}, then the identity $(1-x)^{\alpha+n}(1+x)^{\beta+n} = (1-x)^a(1+x)^b$ would hold where a and b are non-negative whole numbers and $a + b \leqq n - 1$. This implies however that $\alpha + n = a$ which contradicts the requirement $\alpha > -1$.

to all polynomials of H_{n-1}. The degree of $y_n(x)$ can hence not be less than n, since otherwise this polynomial would be orthogonal to itself; this is however impossible, since $y_n(x)$ is not identically zero. Hence $y_n(x)$ is a polynomial of degree n which is orthogonal with weight $(1-x)^\alpha(1+x)^\beta$ to all polynomials of lower degree. As we know, this implies that $y_n(x)$ and $J_n^{(\alpha,\beta)}(x)$ can differ only by a constant factor. Formula (151) had now been proved.

If in the integral (154) we now set $v(x) = y_n(x)$ and again apply formula (155), we find

$$\int_{-1}^{+1}(1-x)^\alpha(1+x)^\beta y_n^2(x)\,dx = (-1)^n\int_{-1}^{+1}u(x)y_n^{(n)}(x)\,dx.$$

But now

$$y_n^{(n)}(x) = n!q_n,$$

where q_n is the leading coefficient of $y_n(x)$. Therefore

$$\int_{-1}^{+1}(1-x)^\alpha(1+x)^\beta y_n^2(x)\,dx = (-1)^n n!q_n\int_{-1}^{+1}(1-x)^{\alpha+n}(1+x)^{\beta+n}dx.$$

Putting $x = 2t - 1$ in the last integral, it assumes the form

$$2^{\alpha+\beta+2n+1}\int_0^1(1-t)^{\alpha+n}t^{\beta+n}\,dt = 2^{\alpha+\beta+2n+1}B(\alpha+n+1,\beta+n+1),$$

where

$$B(p,q) = \int_0^1(1-x)^{p-1}x^{q-1}\,dx$$

is the well-known EULER integral.

Since

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},$$

it follows that

$$\int_{-1}^{+1}(1-x)^\alpha(1+x)^\beta y_n^2(x)\,dx$$

$$= (-1)^n n!q_n 2^{\alpha+\beta+2n+1}\frac{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{\Gamma(\alpha+\beta+2n+2)}. \tag{156}$$

Let us now seek an expression for the leading coefficient q_n of the polynomial $y_n(x)$. In so doing we will also obtain an expression for the coefficient p_n of x^{n-1} in $y_n(x)$ which we shall need subsequently.

To this end we first note that from

$$y_n(x) = (1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}]$$

the LEIBNIZ formula provides the equation

$$y_n(x) = \sum_{k=0}^{n} C_n^k(-1)^k(\alpha+n)\dots(\alpha+n-k+1)(1-x)^{n-k}(\beta+n)\dots(\beta+k+1)(1+x)^k \tag{157}$$

On the other hand, for $x > 1$

$$(x-1)^{-\alpha}(x+1)^{-\beta}\frac{d^n}{dx^n}[(x-1)^{\alpha+n}(x+1)^{\beta+n}] = \sum_{k=0}^{n} C_n^k(\alpha+n)\dots(\alpha+n-k+1)(x-1)^{n-k}(\beta+n)\dots(\beta+k+1)(x+1)^k. \tag{158}$$

A comparison of (157) and (158) shows that for $x > 1$ the polynomial $y_n(x)$ can be represented in the form

$$y_n(x) = (-1)^n(x-1)^{-\alpha}(x+1)^{-\beta}\frac{d^n}{dx^n}[(x-1)^{\alpha+n}(x+1)^{\beta+n}]. \tag{159}$$

From NEWTON's binomial formula

$$(x-1)^{\alpha+n} = x^{\alpha+n}\left(1-\frac{1}{x}\right)^{\alpha+n} = x^{\alpha+n} - (\alpha+n)x^{\alpha+n-1} + \cdots,$$

$$(x+1)^{\beta+n} = x^{\beta+n}\left(1+\frac{1}{x}\right)^{\beta+n} = x^{\beta+n} + (\beta+n)x^{\beta+n-1} + \cdots.$$

From this it follows that

$$(x-1)^{\alpha+n}(x+1)^{\beta+n} = x^{\alpha+\beta+2n} + (\beta-\alpha)x^{\alpha+\beta+2n-1} + \cdots$$

and

$$\frac{d^n}{dx^n}[(x-1)^{\alpha+n}(x+1)^{\beta+n}] = (\alpha+\beta+2n)\dots(\alpha+\beta+n+1)x^{\alpha+\beta+n} + (\beta-\alpha)(\alpha+\beta+2n-1)\dots(\alpha+\beta+n)x^{\alpha+\beta+n-1} + \cdots \tag{160}$$

We find similarly

$$(x-1)^{-\alpha}(x+1)^{-\beta} = x^{-\alpha-\beta}\left(1-\frac{1}{x}\right)^{-\alpha}\left(1+\frac{1}{x}\right)^{-\beta} = x^{-\alpha-\beta} + (\alpha-\beta)x^{-\alpha-\beta-1} + \cdots. \tag{161}$$

From (159), (160), and (161) it follows that

$$y_n(x) = (-1)^n[(\alpha+\beta+2n)\dots(\alpha+\beta+n+1)x^n + (\alpha-\beta)n(\alpha+\beta+2n-1)\dots(\alpha+\beta+n+1)x^{n-1} + \cdots],$$

and hence

$$q_n = (-1)^n(\alpha+\beta+2n)\dots(\alpha+\beta+n+1) = (-1)^n\frac{\Gamma(\alpha+\beta+2n+1)}{\Gamma(\alpha+\beta+n+1)}.$$

$$p_n = (-1)^n(\alpha-\beta)n(\alpha+\beta+2n-1)\dots(\alpha+\beta+n+1)$$
$$= (-1)^n n(\alpha-\beta)\frac{\Gamma(\alpha+\beta+2n)}{\Gamma(\alpha+\beta+n+1)}. \tag{162}$$

Putting q_n into (156) and observing that

$$\frac{\Gamma(\alpha+\beta+2n+1)}{\Gamma(\alpha+\beta+2n+2)} = \frac{1}{\alpha+\beta+2n+1},$$

we obtain

$$\int_{-1}^{+1}(1-x)^\alpha(1+x)^\beta y_n^2(x)\,dx$$
$$= 2^{\alpha+\beta+2n+1}\frac{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{\Gamma(\alpha+\beta+n+1)}\frac{n!}{\alpha+\beta+2n+1}. \tag{163}$$

From the generalized RODRIGUES formula (151)

$$J_n^{(\alpha,\beta)}(x) = K_n y_n(x),$$

hence

$$\tilde{J}_n^{(\alpha,\beta)}(x) = (-1)^n\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+\beta+2n+1)}y_n(x), \tag{164}$$

$$\hat{J}_n^{(\alpha,\beta)}(x) = (-1)^n\sqrt{2^{-(\alpha+\beta+2n+1)}\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}\frac{\alpha+\beta+2n+1}{n!}}\,y_n(x), \tag{165}$$

where

$$y_n(x) = (1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}].$$

These formulas are applicable for all $n > 0$. They are also valid for $n = 0$ if $\alpha+\beta+1 \neq 0$. However, if $n = \alpha+\beta+1 = 0$ they are meaningless. It is immediately clear that

$$y_0(x) = 1,$$

and hence

$$\tilde{J}_0^{(\alpha,\beta)}(x) = y_0(x).$$

Now the substitution $x = 2t-1$ gives

$$\int_{-1}^{+1}(1-x)^\alpha(1+x)^\beta\,dx = 2^{\alpha+\beta+1}\int_0^1(1-t)^\alpha t^\beta\,dt$$
$$= 2^{\alpha+\beta+1}B(\alpha+1,\beta+1) = 2^{\alpha+\beta+1}\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}.$$

Therefore, for $\alpha + \beta + 1 = 0$ we have [2]

$$\hat{J}_0^{(\alpha,\beta)}(x) = \frac{1}{\sqrt{\Gamma(\alpha+1)\Gamma(\beta+1)}}.$$

The JACOBI polynomials which have the factor $K_n = \frac{(-1)^n}{2^n n!}$ will occasionally be of service to us. These we shall denote by $P_n^{(\alpha,\beta)}(x)$. Hence

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}].$$

Formulas (164) and (165) then take the form

$$\tilde{J}_n^{(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+\beta+2n+1)}2^n n! P_n^{(\alpha,\beta)}(x), \tag{166}$$

$$\hat{J}_n^{(\alpha,\beta)}(x) = \sqrt{\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}\frac{\alpha+\beta+2n+1}{2^{\alpha+\beta+1}}n!}\,P_n^{(\alpha,\beta)}(x). \tag{167}$$

Finally we note that the generalized RODRIGUES formula also produces a new representation of the TSHEBYSHEFF polynomials, namely

$$T'_n(x) = K_n\sqrt{1-x^2}\,\frac{d^n(1-x^2)^{n-\frac{1}{2}}}{dx^n}.$$

§2. The Recursion Formula, Generating Function, and Differential Equation

According to the general theory of orthogonal polynomials, three successive JACOBI polynomials are related to one another by the recursion formula

$$\tilde{J}_{n+2}^{(\alpha,\beta)}(x) = (x-\alpha_{n+2})\tilde{J}_{n+1}^{(\alpha,\beta)}(x) - \lambda_{n+1}\tilde{J}_n^{(\alpha,\beta)}(x). \tag{168}$$

The parameter α_{n+2} can be easily found by comparing the coefficients of x^{n+1}. From formulas (164) and (162) after simple rearrangements one obtains

$$\alpha_{n+2} = \frac{\beta^2-\alpha^2}{(\alpha+\beta+2n+2)(\alpha+\beta+2n+4)}. \tag{169}$$

The coefficient λ_{n+1} is found from formulas (73), (164), and (163); one obtains

$$\lambda_{n+1} = \frac{(\alpha+n+1)(\beta+n+1)(\alpha+\beta+n+1)}{(\alpha+\beta+2n+1)(\alpha+\beta+2n+2)^2(\alpha+\beta+2n+3)}4(n+1). \tag{170}$$

For the ultraspherical functions having $\beta = \alpha$ formulas (169) and (170) simplify and become

[2] We use $\Gamma(\alpha+\beta+2) = \Gamma(1) = 1$.

$$\alpha_{n+2} = 0, \quad \lambda_{n+1} = \frac{2\alpha + n + 1}{(2\alpha + 2n + 1)(2\alpha + 2n + 3)}(n + 1).$$

Since $\tilde{J}_0^{(\alpha,\beta)}(x) = 1$ and [3]

$$\tilde{J}_1^{(\alpha,\beta)}(x) = x + \frac{\alpha - \beta}{\alpha + \beta + 2}. \tag{171}$$

formula (168) makes it possible to compute the polynomials $\tilde{J}_n^{(\alpha,\beta)}(x)$ in succession; of course the expressions become more unwieldy from step to step; thus, for example, $\tilde{J}_2^{(\alpha,\beta)}(x)$ has the form

$$\tilde{J}_2^{(\alpha,\beta)}(x) = x^2 + 2\frac{\alpha - \beta}{\alpha + \beta + 4}x + \frac{(\alpha - \beta)^2 - (\alpha + \beta) - 4}{(\alpha + \beta + 3)(\alpha + \beta + 4)}.$$

The *generating function* provides another access to the JACOBI polynomials. Indeed, it can be shown that the polynomials $P_n^{(\alpha,\beta)}(x)$ are the coefficients of t^n in the series expansion of the function

$$J(t, x) = \frac{2^{\alpha+\beta}}{\sqrt{1 - 2tx + t^2}} \cdot (1 - t + \sqrt{1 - 2tx + t^2})^{-\alpha}(1 + t + \sqrt{1 - 2tx + t^2})^{-\beta}.$$

However, computation by this method is also very laborious. The most practical means of computing the JACOBI polynomials is the formula

$$P_n^{(\alpha,\beta)}(x) = \frac{1}{2^n n!}\sum_{k=0}^{n} C_n^k \frac{\Gamma(\alpha + n + 1)}{\Gamma(\alpha + n - k + 1)} \frac{\Gamma(\beta + n + 1)}{\Gamma(\beta + k + 1)} \cdot (x - 1)^{n-k}(x + 1)^k, \tag{172}$$

which is obtained directly from (157)

This formula gives in particular

$$P_n^{(\alpha,\beta)}(1) = \frac{\Gamma(\alpha + n + 1)}{n!\Gamma(\alpha + 1)}, \tag{173}$$

$$P_n^{(\alpha,\beta)}(-1) = (-1)^n \frac{\Gamma(\beta + n + 1)}{n!\Gamma(\beta + 1)}. \tag{174}$$

Theorem. *The polynomial $J_n^{(\alpha,\beta)}(x)$ is a solution of the differential equation*

$$(1 - x^2)y'' + [\beta - \alpha - (\alpha + \beta + 2)x]y' + n(\alpha + \beta + n + 1)y = 0. \tag{175}$$

Proof. We consider

$$y = J_n^{(\alpha,\beta)}(x) = (1 - x)^{-\alpha}(1 + x)^{-\beta}u^{(n)} \tag{176}$$

where

$$u = (1 - x)^{\alpha+n}(1 + x)^{\beta+n}.$$

[3] Equation (171) follows directly from (164) and the generalized RODRIGUES formula.

Since

$$u' = (1 - x)^{\alpha+n-1}(1 + x)^{\beta+n-1}[\beta - \alpha - (\alpha + \beta + 2n)x],$$

we have

$$(1 - x^2)u' = [\beta - \alpha - (\alpha + \beta + 2n)x]u.$$

We differentiate both sides of this equation $(n + 1)$ times. Using the LEIBNIZ formula we then obtain

$$\sum_{k=0}^{n+1} C_{n+1}^k (1 - x^2)^{(k)} u^{(n+2-k)} = \sum_{k=0}^{n+1} C_{n+1}^k [\beta - \alpha - (\alpha + \beta + 2n)x]^{(k)} u^{(n+1-k)}.$$

From this it follows that

$$\begin{aligned}(1 - x^2)u^{(n+2)} - 2(n + 1)xu^{(n+1)} - (n + 1)nu^{(n)} \\ = [\beta - \alpha - (\alpha + \beta + 2n)x]u^{(n+1)} - (n + 1)(\alpha + \beta + 2n)u^{(n)},\end{aligned}$$

and hence

$$\begin{aligned}(1 - x^2)u^{(n+2)} + [\alpha - \beta + (\alpha + \beta - 2)x]u^{(n+1)} \\ + (n + 1)(\alpha + \beta + n)u^{(n)} = 0.\end{aligned} \tag{177}$$

But from (176) it follows that

$$u^{(n)} = (1 - x)^{\alpha}(1 + x)^{\beta}y.$$

We now compute $u^{(n+1)}$ and $u^{(n+2)}$, put the expressions so obtained into (177), and thus arrive at (175).

§3. Estimates of the JACOBI Polynomials. The Expansion Problem

In seeking certain estimates for the JACOBI polynomials we shall here restrict ourselves to cases in which

$$\sigma = \max\{\alpha, \beta\} \geqq -\tfrac{1}{2}, \tag{178}$$

since for $\sigma < -2^{-1}$ the investigations are substantially more difficult and involved.

Theorem 1. *Under the condition* (178) *the absolute value* $|J_n^{(\alpha,\beta)}(x)|$ *assumes its greatest value (for* $x \in [-1, +1]$*) at one of the two points* ± 1.[4]

Proof. We denote $J_n^{(\alpha,\beta)}(x)$ by y and put

$$\psi(x) = y^2 + \frac{1 - x^2}{n(\alpha + \beta + n + 1)} y'^2. \tag{179}$$

[4] For the case $\sigma < -\frac{1}{2}$ however this greatest value is assumed in the interior of the interval $(-1, +1)$.

Then

$$\psi'(x) = 2yy' - \frac{2x}{n(\alpha+\beta+n+1)}y'^2 + \frac{1-x^2}{n(\alpha+\beta+n+1)}2y'y''.$$

Since y is a solution of the differential equation (175),

$$\psi'(x) = \frac{2}{n(\alpha+\beta+n+1)}[(\alpha+\beta+1)x - (\beta-\alpha)]y'^2.$$

First assume

$$\alpha > -\tfrac{1}{2}, \quad \beta > -\tfrac{1}{2}. \tag{180}$$

Then $\alpha+\beta+1>0$ and

$$\psi'(x) = K^2(x-x_0)y'^2$$

where

$$x_0 = \frac{\beta-\alpha}{\alpha+\beta+1}.$$

From (180)

$$-1 < x_0 < 1,$$

and the last expression for $\psi'(x)$ implies that $\psi'(x) < 0$ for $-1 \leqq x \leqq x_0$ while $\psi'(x) > 0$ for $x_0 < x \leqq 1$, such that $\psi(x)$ decreases on the interval $[-1, x_0]$ and increases on the interval $[x_0, +1]$. For $x_0 \leqq x < +1$ therefore

$$\psi(x) < \psi(1).$$

But now

$$[J_n^{(\alpha,\beta)}(x)]^2 \leqq \psi(x), \qquad [J_n^{(\alpha,\beta)}(1)]^2 = \psi(1),$$

such that we obtain

$$|J_n^{(\alpha,\beta)}(x)| < |J_n^{(\alpha,\beta)}(1)| \qquad (x_0 \leqq x < 1).$$

Therefore, on the interval $[x_0, 1]$ the absolute value of the polynomial $J_n^{(\alpha,\beta)}(x)$ attains its greatest value at the point $x = +1$.

Similarly we find

$$\max_{-1\leqq x\leqq x_0} |J_n^{(\alpha,\beta)}(x)| = |J_n^{(\alpha,\beta)}(-1)|.$$

The theorem is herewith proved for the case (180).

Now suppose

$$\alpha \geqq -\tfrac{1}{2}, \quad \beta < -\tfrac{1}{2}. \tag{181}$$

Then for the sum $\alpha+\beta+1$ there are three possibilities:

$$\alpha+\beta+1 \begin{cases} = 0, \\ > 0, \\ < 0. \end{cases}$$

In the first case $\psi'(x) > 0$ for all real values of x with the exception of a finite number of points at which $y' = 0$. In the second case it follows from $2\beta < -1$ that $\psi'(x) > 0$ for $-1 \leqq x \leqq 1$ (again with the exception of the zeros of y'). In the third case we observe that $2\alpha \geqq -1$ and again obtain that $\psi'(x) > 0$ everywhere on $[-1, +1]$, excepting the zeros of y' (and the point $+1$ if $\alpha = -\frac{1}{2}$). Under condition (181) therefore the function $\psi(x)$ is monotone increasing (in the strict sense) and the same considerations as above lead to the equation

$$\max_{-1 \leqq x \leqq 1} |J_n^{(\alpha,\beta)}(x)| = |J_n^{(\alpha,\beta)}(+1)|.$$

If now $\alpha < -\frac{1}{2}$, $\beta \geqq -\frac{1}{2}$, then entirely analogous considerations lead to

$$\max_{-1 \leqq x \leqq 1} J_n^{(\alpha,\beta)}(x)| = |J_n^{(\alpha,\beta)}(-1)|.$$

If $\alpha > -\frac{1}{2}$, $\beta = -\frac{1}{2}$, then $\psi'(x) = K^2(x+1)y'^2$, and hence $\psi'(x) > 0$ for $x > -1$. Therefore $\psi(x)$ is increasing and $|J_n^{(\alpha,\beta)}(x)|$ attains its maximum value at $x = +1$. The case $\alpha = -\frac{1}{2}$, $\beta > -\frac{1}{2}$ is settled similarly.

Hence only the case $\alpha = \beta = -\frac{1}{2}$ remains (the case of the TCHEBYSHEFF polynomials). Here however $\psi'(x) = 0$, so that $\psi(x)$ is a constant. Now

$$[J_n^{(\alpha,\beta)}(x)]^2 \leqq \psi(x) = \psi(1) = [J_n^{(\alpha,\beta)}(1)]^2,$$

wherewith the theorem is finally proved.

Let us now note certain simple properties of the function

$$\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1}\, dt \qquad (x > 0).$$

Its derivative $\Gamma'(x)$ has a single root u_0 with $1 < u_0 < 2$.

For since

$$\Gamma''(x) = \int_0^{+\infty} e^{-t} t^{x-1} \ln^2 t\, dt > 0,$$

$\Gamma'(x)$ is increasing and may therefore have at most one real root. On the other hand

$$\Gamma(1) = \Gamma(2) = 1,$$

such that the mean value theorem implies that there is one root of $\Gamma'(x)$ between 1 and 2. For $x \geqq u_0$ the function $\Gamma(x)$ obviously increases; we now estimate this rate of increase.

Lemma. *The estimate*

$$\frac{\Gamma(\alpha + \beta + n + 1)}{\Gamma(\alpha + n + 1)} < Cn^{\beta},$$

where C is a number depending on α and β, is valid for $\alpha > -1$, $\beta > -1$, and every positive integer n.

Proof. We have

$$\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)} = \frac{(\alpha+\beta+n)(\alpha+\beta+n-1)\cdots(\alpha+\beta+2)\Gamma(\alpha+\beta+2)}{(\alpha+n)(\alpha+n-1)\cdots(\alpha+2)\Gamma(\alpha+2)},$$

hence

$$\ln\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)} = \ln\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)} + \sum_{k=2}^{n}\ln\left(1+\frac{\beta}{\alpha+k}\right).$$

For $x > -1$ however

$$\ln(1+x) \leqq x,$$

and therefore

$$\ln\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)} \leqq \ln\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)} + \beta\sum_{k=2}^{n}\frac{1}{\alpha+k}.$$

On the other hand, for $k-1 < x < k$

$$\frac{1}{\alpha+k} < \frac{1}{\alpha+x} < \frac{1}{\alpha+k-1} \qquad (k \geqq 2),$$

hence

$$\frac{1}{\alpha+k} < \int_{k-1}^{k}\frac{dx}{\alpha+x} < \frac{1}{\alpha+k-1} \qquad (k \geqq 2)$$

and therefore

$$\sum_{k=2}^{n}\frac{1}{\alpha+k} < \ln\frac{\alpha+n}{\alpha+1} < \sum_{k=1}^{n-1}\frac{1}{\alpha+k}.$$

From this it follows that

$$\ln\frac{\alpha+n}{\alpha+1} - \frac{1}{\alpha+1} < \sum_{k=2}^{n}\frac{1}{\alpha+k} < \ln\frac{\alpha+n}{\alpha+1}$$

and finally

$$\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)} \leqq \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)}\left(\frac{\alpha+n}{\alpha+1}\right)^{\beta} \qquad (\text{for } \beta \geqq 0).$$

$$\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)} < \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)}\,e^{\frac{-\beta}{\alpha+1}}\left(\frac{\alpha+n}{\alpha+1}\right)^{\beta} \qquad (\text{for } \beta < 0).$$

This completes the proof of the lemma.[5]

Theorem 2. *Under condition* (178) *the normalized* JACOBI *polynomials satisfy the inequality*

$$|\hat{J}_n^{(\alpha,\beta)}(x)| < Mn^{\sigma+\frac{1}{2}}$$
$$(-1 \leqq x \leqq 1, \quad n = 1, 2, 3, \ldots), \tag{182}$$

where M is a constant depending on α and β.

Proof. From the first theorem

$$|\hat{J}_n^{(\alpha,\beta)}(x)| \leqq \max\{|\hat{J}_n^{(\alpha,\beta)}(+1)|, \quad |\hat{J}_n^{(\alpha,\beta)}(-1)|\}.$$

But according to (167), (173), and (174)

$$|\hat{J}_n^{(\alpha,\beta)}(+1)| = \frac{1}{\Gamma(\alpha+1)}\sqrt{\frac{\alpha+\beta+2n+1}{2^{\alpha+\beta+1}}}\sqrt{\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\beta+n+1)}\frac{\Gamma(\alpha+n+1)}{\Gamma(n+1)}},$$

$$|\hat{J}_n^{(\alpha,\beta)}(-1)| = \frac{1}{\Gamma(\beta+1)}\sqrt{\frac{\alpha+\beta+2n+1}{2^{\alpha+\beta+1}}}\sqrt{\frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)}\frac{\Gamma(\beta+n+1)}{\Gamma(n+1)}}.$$

The lemma now gives

$$|\hat{J}_n^{(\alpha,\beta)}(+1)| < C_1 n^{\alpha+\frac{1}{2}}, \qquad |\hat{J}_n^{(\alpha,\beta)}(-1)| < C_2 n^{\beta+\frac{1}{2}},$$

where C_1 and C_2 are constants depending on α and β. The theorem is herewith proved.

Theorem 3. *Let $\sigma \geqq -\frac{1}{2}$, and let p be a positive integer which is not less than $2\sigma + 2$. Then on the interval $[-1, +1]$ every function $f(x)$ with a continuous pth derivative can be expanded in a uniformly convergent* FOURIER *series of polynomials $\hat{J}_n^{(\alpha,\beta)}(x)$.*

Proof. From (182)

$$|K_n(t,x)| \leqq \sum_{k=0}^{n} |\hat{J}_k^{(\alpha,\beta)}(t)|\,|\hat{J}_k^{(\alpha,\beta)}(x)| \leqq A^2 + M^2 \sum_{k=1}^{n} k^{2\sigma+1},$$

where $A = \hat{J}_0^{(\alpha,\beta)}(x)$. For $n \geqq 1$ therefore

$$|K_n(t,x)| < M_1 n^{2\sigma+2},$$

where M_1 is another constant. From this we obtain for the LEBESGUE function the estimate

$$L_n(x) < M_2 n^{2\sigma+2}.$$

[5] For the ratio $\dfrac{\left(\dfrac{\alpha+n}{\alpha+1}\right)^{\beta}}{n^{\beta}}$ tends to $\dfrac{1}{(\alpha+1)^{\beta}}$ with increasing n and is therefore bounded.

But then from the JACKSON theorem

$$\lim_{n\to\infty} n^p E_n(f) = 0,$$

wherewith according to Theorem 7, § 4, Chapter IV our assertion is proved.

The reader can find more information on the JACOBI polynomials in the monograph by SZEGÖ [1].

§4. TCHEBYSHEFF Polynomials of the Second Kind

Those polynomials $U_n(x)$ which form an orthogonal system with weight $\sqrt{1-x^2}$ on the interval $[-1, +1]$ are called TCHEBYSHEFF *polynomials of the second kind.*

Lemma 1. *The identity*

$$\frac{\sin(n+1)\theta}{\sin\theta} = 2^n \cos^n\theta + \sum_{k=0}^{n-1} \lambda_k^{(n)} \cos^k\theta.$$

holds.

Proof. The proof proceeds inductively on the basis of the formulas

$$\sin(n+1)\theta = \sin n\theta \cos\theta + \cos n\theta \sin\theta$$

and

$$\cos n\theta = 2^{n-1} \cos^n\theta + \sum_{k=0}^{n-1} \mu_k^{(n)} \cos^k\theta.$$

Corollary. *For* $|x| \leqq 1$ *the function*

$$\frac{\sin(n+1) \arccos x}{\sqrt{1-x^2}} \tag{183}$$

is a polynomial of degree n with leading coefficient 2^n.

Lemma 2. *The polynomials* $\dfrac{\sin(n+1)\arccos x}{\sqrt{1-x^2}}$ *with weight* $\sqrt{1-x^2}$ *form an orthogonal system on* $[-1, +1]$.

Proof. With the substitution $x = \cos\theta$ the integral

$$\int_{-1}^{+1} \frac{\sin(n+1)\arccos x}{\sqrt{1-x^2}} \frac{\sin(m+1)\arccos x}{\sqrt{1-x^2}} \sqrt{1-x^2}\, dx$$

becomes

$$\int_0^{\pi} \sin(n+1)\theta \sin(m+1)\theta\, d\theta = 0.$$

We therefore obtain all the polynomials $U_n(x)$ from the formula

$$U_n(x) = K_n \frac{\sin (n+1) \arccos x}{\sqrt{1-x^2}}. \tag{184}$$

Remark. Formula (184) can also be derived directly from the definition of the polynomials $U_n(x)$. For every polynomial $R(x) \in H_{n-1}$

$$\int_{-1}^{+1} \sqrt{1-x^2}\, R(x) U_n(x)\, dx = 0.$$

If we here make the substitution $x = \cos\theta$ we find

$$\int_0^{\pi} R(\cos\theta) U_n(\cos\theta) \sin^2\theta\, d\theta = 0.$$

Now since $\sin^2\theta U_n(\cos\theta)$ is an even trigonometric polynomial of order $n+2$,

$$\sin^2\theta U_n(\cos\theta) = \sum_{k=0}^{n+2} A_k \cos k\theta.$$

If we substitute this expression in the preceding integral and take for $R(\cos\theta)$ the function $\cos m\theta$ with $m < n$ we obtain

$$\int_0^{\pi} \cos m\theta \left[\sum_{k=0}^{n+2} A_k \cos k\theta\right] d\theta = 0.$$

From this it follows that $A_m = 0$. Hence

$$\sin^2\theta U_n(\cos\theta) = A_n \cos n\theta + A_{n+1} \cos (n+1)\theta + A_{n+2} \cos (n+2)\theta.$$

If we here put $\theta = 0$ and $\theta = \pi$ we obtain two conditions for the coefficients A:

$$A_n + A_{n+1} + A_{n+2} = 0, \quad A_n - A_{n+1} + A_{n+2} = 0,$$

whence it follows that $A_{n+1} = 0$ and $A_{n+2} = -A_n$. Therefore

$$\sin^2\theta U_n(\cos\theta) = A_n[\cos n\theta - \cos (n+2)\theta] = 2A_n \sin (n+1)\theta \sin\theta,$$

which likewise leads to formula (184).

We obtain the polynomial $\tilde{U}_n(x)$ with leading coefficient 1 upon multiplication by $K_n = 2^{-n}$. In view of the identity

$$\int_{-1}^{+1} U_n^2(x)\sqrt{1-x^2}\, dx = K_n^2 \int_0^{\pi} \sin^2 (n+1)\theta\, d\theta = \frac{\pi}{2} K_n^2$$

the normalized polynomials $\hat{U}_n(x)$ are clearly obtained for $K_n = \sqrt{\frac{2}{\pi}}$.

The recursion formula for the $U_n(x)$ is most easily obtained from the formula

$$\sin (n+3)\theta + \sin (n+1)\theta = 2 \sin (n+2)\theta \cos \theta,$$

from which

$$\tilde{U}_{n+2}(x) = x\tilde{U}_{n+1}(x) - \tfrac{1}{4}\tilde{U}_n(x) \tag{185}$$

follows.

Since

$$\tilde{U}_0(x) = 1, \quad \tilde{U}_1(x) = x,$$

it follows that

$$\begin{aligned} \tilde{U}_2(x) &= x^2 - \tfrac{1}{4}, \\ \tilde{U}_3(x) &= x^3 - \tfrac{1}{2}x, \\ \tilde{U}_4(x) &= x^4 - \tfrac{3}{4}x + \tfrac{1}{16}. \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

In order to be able to write down the continued fraction corresponding to $U_n(x)$, we note that

$$\alpha_{n+2} = 0, \quad \lambda_{n+1} = \tfrac{1}{4} \qquad (n = 0, 1, 2, \ldots).$$

Moreover [6] $\alpha_1 = 0$ and $\lambda_0 = \dfrac{\pi}{2}$. Since now for $x > 1$

$$\int_{-1}^{+1} \frac{\sqrt{1-t^2}}{x-t}\,dt = \pi(x - \sqrt{x^2-1}),$$

we obtain

$$2(x - \sqrt{x^2-1}) = \cfrac{1}{x - \cfrac{\frac{1}{4}}{x - \cfrac{\frac{1}{4}}{x - \cdots}}}.$$

We also note that the roots of the polynomials $U_n(x)$ are the numbers

$$x_k^{(n)} = \cos \frac{k\pi}{n+1} \qquad (k = 1, 2, \ldots, n).$$

From this for $n \to \infty$ the same limit function for the density of distribution of roots is obtained as for the polynomials $T_n(x)$, namely

$$\frac{1}{\sqrt{1-x^2}}.$$

[6] Since α_1 is the root of $U_1(x)$ and $\lambda_0 = \displaystyle\int_{-1}^{+1} \sqrt{1-x^2}\,dx$.

This is already clear from the fact that the polynomials $U_n(x)$ are derivatives [7] of the polynomials $T_{n+1}(x)$, since

$$\frac{dT_{n+1}(x)}{dx} = \frac{d}{dx}[\cos(n+1)\arccos x] = (n+1)\frac{\sin(n+1)\arccos x}{\sqrt{1-x^2}}.$$

We shall now consider the expansion in terms of the polynomials $U_n(x)$. We note first of all that

$$|\hat{U}_n(x)| \leqq \sqrt{\frac{2}{\pi}}(n+1) \qquad (-1 \leqq x \leqq 1),$$

$$|\hat{U}_n(x)| \leqq \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{1-x^2}} \qquad (-1 < x < 1).$$

On the basis of the general theorems of Chapter IV now follow:

I. *Every function with a continuous third derivative can be expanded in a uniformly convergent* FOURIER *series of polynomials* $\hat{U}_n(x)$.

II. *Every function of* $L^2_{\sqrt{1-x^2}}$ *can be expanded in a* FOURIER *series of polynomials* $\hat{U}_n(x)$ *at each point* x *for which*

$$\int_{-1}^{+1}\left[\frac{f(t)-f(x)}{t-x}\right]^2\sqrt{1-t^2}\,dt < +\infty$$

holds.

With the help of trigonometric FOURIER series we obtain the more precise

Theorem 1. *If a function defined on the interval* $[-1, +1]$ *satisfies the* DINI-LIPSCHITZ *condition*

$$\lim_{\delta \to 0}\omega(\delta)\ln\delta = 0,$$

then it can be expanded in a FOURIER *series of polynomials* $\hat{U}_n(x)$ *in the interval* $(-1, +1)$ *which converges uniformly on every interval* $[-1+h, 1-h]$.

Proof. The function $\sin\theta f(\cos\theta)$ also satisfies the DINI-LIPSCHITZ condition [8] and can hence be expanded in a uniformly convergent trigonometric

[7] Related to this is the general problem of studying orthogonal systems (with an arbitrary weight) of such functions whose derivatives again form an orthogonal system (with some other weight). For this see B. M. GAGAEV [1].

[8] If $|\theta_1 - \theta_2| \leqq \delta$ and if we set $f(\cos\theta) = \psi(\theta)$, then we have
$|\sin\theta_1\psi(\theta_1) - \sin\theta_2\psi(\theta_2)| \leqq |\sin\theta_1 - \sin\theta_2|\,|\psi(\theta_1)| + |\sin\theta_2|\,|\psi(\theta_1) - \psi(\theta_2)| \leqq K\delta + \omega(\delta)$.

FOURIER series. Since this function is odd,

$$\sin\theta f(\cos\theta) = \sum_{n=0}^{\infty} b_{n+1}\sin(n+1)\theta$$

$$\left(b_n = \frac{2}{\pi}\int_0^{\pi}\sin\theta f(\cos\theta)\sin n\theta\, d\theta\right).$$

From this for $0 < \theta < \pi$ the expansion

$$f(\cos\theta) = \sum_{n=0}^{\infty} b_{n+1}\frac{\sin(n+1)\theta}{\sin\theta}$$

follows, whereby convergence is uniform in every interval $[\eta, \pi - \eta]$. Now putting $\theta = \arccos x$ we obtain the theorem, since

$$b_{n+1} = \sqrt{\frac{2}{\pi}}\int_{-1}^{+1} f(x)\hat{U}_n(x)\sqrt{1-x^2}\,dx.$$

Finally we prove an interesting extremal property of the polynomials $U_n(x)$.

Theorem 2. (E. I. ZOLOTAREV and A. N. KORKIN). *If for all polynomials $\tilde{P}_n(x)$ of degree n with leading coefficient equal to one we form the integral*

$$\int_{-1}^{+1} |\tilde{P}_n(x)|\,dx, \tag{186}$$

then the polynomial $\tilde{U}_n(x)$ yields the minimum value of this integral.[9]

The proof requires certain preliminaries.

Lemma 1. *Let n by a positive integer and m be one of the numbers 0, 1, 2, ..., n. Then*

$$\sum_{k=1}^{n}(-1)^k\cos\frac{km\pi}{n+1} = \frac{(-1)^{n+m}-1}{2}. \tag{187}$$

Proof.

$$\sum_{k=1}^{n}(-1)^k\cos\frac{km\pi}{n+1} = R\left[\sum_{k=1}^{n}\left(-e^{\frac{im\pi}{n+1}}\right)^k\right] = R\left[\frac{(-1)^{n+m}-e^{\frac{im\pi}{n+1}}}{1+e^{\frac{im\pi}{n+1}}}\right].$$

[9] A. N. KORKIN and E. I. ZOLOTAREV [1].

If one of the two numbers n and m is even and the other odd, then

$$\frac{(-1)^{n+m} - e^{\frac{im\pi}{n+1}}}{1 + e^{\frac{im\pi}{n+1}}} = -1. \tag{188}$$

but if n and m are both either odd or even, then the left side of (188) is pure imaginary, since

$$\frac{1 - e^{i\theta}}{1 + e^{i\theta}} = \frac{1 - \cos\theta - i\sin\theta}{1 + \cos\theta + i\sin\theta} = -i\,\mathrm{tg}\,\frac{\theta}{2}.$$

The lemma is herewith proved.

Lemma 2. *If n is a positive integer and r is one of the numbers* 0, 1, . . . , $n - 1$, *then*

$$I = \int_0^{\pi} \cos^k\theta \sin\theta \,\mathrm{sign}\,[\sin(n+1)\theta]\,d\theta = 0.$$

Proof. For

$$k\pi < (n+1)\theta < (k+1)\pi$$

we have

$$\mathrm{sign}\,[\sin(n+1)\theta] = (-1)^k.$$

Whence it follows that

$$I = \sum_{k=0}^{n} (-1)^k \int_{\frac{k\pi}{n+1}}^{\frac{(k+1)\pi}{n+1}} \cos^r\theta \sin\theta\,d\theta,$$

and therefore

$$I = \frac{1}{r+1} \sum_{k=0}^{n} (-1)^{k+1} \left[\cos^{r+1}\frac{(k+1)\pi}{n+1} - \cos^{r+1}\frac{k\pi}{n+1} \right].$$

But now

$$\cos^{r+1}\theta = \sum_{m=0}^{r+1} a_m \cos m\theta,$$

wherewith everything reduces to proving the equation

$$\sum_{k=0}^{n} (-1)^{k+1} \left[\cos \frac{(k+1)m\pi}{n+1} - \cos \frac{km\pi}{n+1} \right] = 0. \tag{189}$$

This latter may be rewritten

$$\sum_{k=1}^{n+1} (-1)^k \cos \frac{km\pi}{n+1} + \sum_{k=0}^{n} (-1)^k \cos \frac{km\pi}{n+1} = 0$$

or

$$(-1)^{n+1} \cos m\pi + 2 \sum_{k=1}^{n} (-1)^k \cos \frac{km\pi}{n+1} + 1 = 0,$$

an equation which is equivalent to (187).

Corollary. *If n is a positive integer and r one of the numbers* 0, 1, 2, . . . , $n - 1$, *then*

$$\int_{-1}^{+1} x^r \operatorname{sign} [\tilde{U}_n(x)]\, dx = 0. \tag{190}$$

Lemma 3. *The equation*

$$\int_{-1}^{+1} x^n \operatorname{sign} [\tilde{U}_n(x)]\, dx = \frac{1}{2^{n-1}} \tag{191}$$

holds.

Proof. Similarly as for the preceding lemma, the proof reduces to showing that the equation

$$\int_0^{\pi} \cos^n \theta \sin \theta \operatorname{sign} [\sin (n+1)\theta]\, d\theta = \frac{1}{2^{n-1}}$$

holds. This integral is equal to

$$\frac{1}{n+1} \sum_{k=0}^{n} (-1)^{k+1} \left[\cos^{n+1} \frac{(k+1)\pi}{n+1} - \cos^{n+1} \frac{k\pi}{n+1} \right].$$

Observing that

$$\cos^{n+1} \theta = \frac{\cos (n+1)\theta}{2^n} + \sum_{m=0}^{n} a_m \cos m\theta$$

and using (189), the lemma reduces to the identity

$$\frac{1}{2^n(n+1)} \sum_{k=0}^{n} (-1)^{k+1}[\cos(k+1)\pi - \cos k\pi] = \frac{1}{2^{n-1}}$$

which is obviously correct.

We now return to the proof of the ZOLATAREV-KORKIN theorem.

Proof of Theorem 2. Let $\tilde{P}_n(x)$ be any polynomial of degree n with leading coefficient equal to one. From (190) and (191) we then have

$$\int_{-1}^{+1} \tilde{P}_n(r) \operatorname{sign}[\tilde{U}_n(x)]\, dx = \frac{1}{2^{n-1}},$$

whence

$$\frac{1}{2^{n-1}} \leqq \int_{-1}^{+1} |\tilde{P}_n(x)|\, dx.$$

However, for $\tilde{P}_n(x) = \tilde{U}_n(x)$ the equality sign holds; thus the integral (186) is actually minimized by the polynomial $\tilde{U}_n(x)$. It remains only to show that the problem has no other solution. Suppose

$$\int_{-1}^{+1} |\tilde{P}_n(x)|\, dx = \frac{1}{2^{n-1}},$$

then

$$\int_{-1}^{+1} |\tilde{P}_n(x)|\{1 - \lambda(x)\}\, dx = 0,$$

where $\lambda(x) = \operatorname{sign}[\tilde{P}_n(x)] \operatorname{sign}[\tilde{U}_n(x)]$. Therefore $\lambda(x) = 1$; the polynomials $\tilde{P}_n(x)$ and $\tilde{U}_n(x)$ therefore have the same sign. Expressed in another manner: the roots of both polynomials are the same and in addition both have leading coefficient one; they are therefore identical.

The ZOLATAREV-KORKIN theorem has been repeatedly generalized in various directions [9]; the interested reader is referred to the literature cited below.

§5. The JACOBI Polynomials for $\alpha = \frac{1}{2}$, $\beta = -\frac{1}{2}$

In concluding this chapter on JACOBI polynomials we consider the special case of the polynomials $W_n(x)$ which form an orthogonal system on $[-1, +1]$ with weight

$$p(x) = \sqrt{\frac{1-x}{1+x}}.$$

[9] N. J. AKHIEZER and M. G. KREIN [2], J. L. GERONIMUS [2, 3].

An explicit formula for the polynomials $W_n(x)$ can be obtained in the following way: for every polynomial $R(x) \in H_{n-1}$

$$\int_{-1}^{+1} \sqrt{\frac{1-x}{1+x}}\, R(x) W_n(x)\, dx = 0.$$

Substituting $x = \cos\theta$ and observing that

$$\sqrt{\frac{1-\cos\theta}{1+\cos\theta}} = \tan\frac{\theta}{2},$$

we obtain

$$\int_0^{\pi} R(\cos\theta)(1-\cos\theta) W_n(\cos\theta)\, d\theta = 0.$$

But now $W_n(\cos\theta)$ is an even trigonometric polynomial of order n, and therefore

$$(1-\cos\theta) W_n(\theta) = \sum_{k=0}^{n+1} A_k \cos k\theta.$$

Substituting this expression into the preceding integral and taking for $R(\cos\theta)$ the function $\cos m\theta$ with $m < n$, we find $A_m = 0$. Therefore

$$(1-\cos\theta) W_n(\cos\theta) = A_n \cos n\theta + A_{n+1} \cos(n+1)\theta.$$

For $\theta = 0$ it follows from this that $A_{n+1} = -A_n$, and therefore

$$(1-\cos\theta) W_n(\cos\theta) = A_n[\cos n\theta - \cos(n+1)\theta].$$

Hence [11]

$$W_n(\cos\theta) = A_n \frac{\sin\frac{2n+1}{2}\theta}{\sin\frac{\theta}{2}}.$$

[11] This formula can be obtained still more simply from the identity

$$\sqrt{\frac{1-x}{1+x}} = \frac{1}{\sqrt{1-x^2}}(x-1),$$

so that formula (111) can be used. For with the notation of § 5, Chapter IV we have

$$Q_n(x) = \begin{vmatrix} T_n(1) & T_n(x) \\ T_{n+1}(1) & T_{n+1}(x) \end{vmatrix} = T_{n+1}(x) - T_n(x),$$

whence

$$W_n(x) = K_n \frac{T_{n+1}(x) - T_n(x)}{1-x} = K_n \frac{\cos(n+1)\theta - \cos n\theta}{1-\cos\theta} = -K_n \frac{\sin\frac{2n+1}{2}\theta}{\sin\frac{\theta}{2}}.$$

With the help of formula (175) of Volume I we can write the last equation in the form

$$W_n(\cos\theta) = A_n[1 + 2(\cos\theta + \cos 2\theta + \cdots + \cos n\theta)];$$

from this we obtain a representation of W_n in terms of TCHEBYSHEFF polynomials:

$$W_n(x) = A_n\{1 + 2[T_1(x) + T_2(x) + \cdots + T_n(x)]\}.$$

Since the leading coefficient of $T_n(x)$ is equal to 2^{n-1},

$$\tilde{W}_n(x) = \frac{1}{2^n}\,\frac{\sin\dfrac{2n+1}{2}\theta}{\sin\dfrac{\theta}{2}} \qquad (\theta = \arccos x).$$

In order to find the normalized polynomials $\hat{W}_n(x)$, we compute

$$\int_{-1}^{+1}\sqrt{\frac{1-x}{1+x}}\,W_n^2(x)\,dx = A_n^2\int_0^{\pi}(1-\cos\theta)\,\frac{\sin^2\dfrac{2n+1}{2}\theta}{\sin^2\dfrac{\theta}{2}}\,d\theta$$

$$= 2A_n^2\int_0^{\pi}\sin^2\frac{2n+1}{2}\theta\,d\theta = A_n^2\pi.$$

Hence

$$\hat{W}_n(x) = \frac{1}{\sqrt{\pi}}\,\frac{\sin\dfrac{2n+1}{2}\theta}{\sin\dfrac{\theta}{2}} \qquad (\theta = \arccos x).$$

From this we obtain the estimates

$$|\hat{W}_n(x)| \leqq \frac{2n+1}{\sqrt{\pi}} \qquad (-1 \leqq x \leqq 1),$$

$$|\hat{W}_n(x)| < \sqrt{\frac{2}{\pi}}\,\frac{1}{\sqrt{1-x}} \qquad (-1 \leqq x < 1).$$

On the basis of these estimates and the general theorems of Chapter IV we can now make assertions with respect to expansion in terms of the polynomials $\hat{W}_n(x)$. For example, every function having a continuous third derivative on the interval $[-1, +1]$ can be expanded into a uniformly con-

vergent FOURIER series of polynomials $\hat{W}_n(x)$. A still better result can be obtained by observing that

$$K_n(t, x) = \sum_{k=0}^{n} \hat{W}_k(t)\hat{W}_k(x)$$

$$= \frac{1}{2\pi \sin\frac{\theta}{2}\sin\frac{\tau}{2}} \sum_{k=0}^{n} \left[\cos\frac{2k+1}{2}(\tau-\theta) - \cos\frac{2k+1}{2}(\tau+\theta)\right]$$

where $\theta = \arccos x$ and $\tau = \arccos t$. For from this it follows that

$$K_n(t, x) = \frac{1}{4\pi \sin\frac{\theta}{2}\sin\frac{\tau}{2}} \left[\frac{\sin(n+1)(\tau-\theta)}{\sin\frac{\tau-\theta}{2}} - \frac{\sin(n+1)(\tau+\theta)}{\sin\frac{\tau+\theta}{2}}\right],$$

and therefore

$$L_n(x) = \int_{-1}^{+1} \sqrt{\frac{1-t}{1+t}}\,|K_n(t, x)|\,dt$$

$$= \frac{1}{2\pi \sin\frac{\theta}{2}} \int_0^{\pi} \sin\frac{\tau}{2} \left|\frac{\sin(n+1)(\tau-\theta)}{\sin\frac{\tau-\theta}{2}} - \frac{\sin(n+1)(\tau+\theta)}{\sin\frac{\tau+\theta}{2}}\right| d\tau.$$

From this we obtain by the usual methods

$$L_n(x) < \frac{C \ln n}{\sin\frac{\theta}{2}} = \frac{C_1 \ln n}{\sqrt{1-x}};$$

Every function $f(x)$ satisfying the DINI-LIPSCHITZ condition $\lim_{\delta\to 0} \omega(\delta) \ln \delta = 0$ can therefore be expanded at all points of the half-open interval $[-1, +1)$ in a FOURIER series of polynomials $\hat{W}_n(x)$ which converges uniformly on every interval $[-1, 1-h]$.

From the formula

$$W_n(x) = A_n \frac{\sin\frac{2n+1}{2}\theta}{\sin\frac{\theta}{2}}$$

it follows that the roots x_k of the polynomial $W_n(x)$ are the points

$$x_k = \cos\frac{2k\pi}{2n+1} \qquad (k = 1, 2, \ldots, n).$$

We recommend that the reader find the recursion formula for the polynomials $W_n(x)$ and in conjunction with this the expansion of the integral

$$\int_{-1}^{+1} \sqrt{\frac{1-t}{1+t}} \frac{dt}{x-t} = \pi \frac{\sqrt{x+1} - \sqrt{x-1}}{\sqrt{x+1}} \qquad (x > 1)$$

in a continued fraction.

CHAPTER VII

THE MOMENT PROBLEM FOR FINITE INTERVALS

§1. Formulation of the Problem

As we have already seen, the numbers

$$\mu_n = \int_a^b x^n p(x)\,dx \qquad (n = 0, 1, 2, \ldots) \quad (192)$$

are called the *moments* of the weight function. Knowing these numbers it is possible to construct an orthogonal system of polynomials with the weight $p(x)$ without knowing this weight itself. This is almost self-evident, since a weight function is uniquely determined by its moments (if, as heretofore, equivalent functions are not considered to be distinct). If $p(x)$ is *square-summable* this follows from the fact that the power system $\{x^n\}$ is complete in L^2; however, there is also the much more general

Theorem 1. *If the function $f(x)$ is summable and if*

$$\int_a^b x^n f(x)\,dx = 0 \qquad (n = 0, 1, 2, \ldots), \quad (193)$$

then $f(x)$ is identically zero.

Proof. We define

$$F(x) = \int_a^x f(t)\,dt$$

and obtain from (193) for $n = 0$ the equation $F(b) = 0$. For $n > 0$ integration by parts gives

$$\int_a^b x^n f(x)\,dx = [x^n F(x)]_a^b - n\int_a^b x^{n-1} F(x)\,dx,$$

from which together with (193) we obtain

$$\int_a^b x^{n-1} F(x)\,dx = 0 \qquad (n = 1, 2, 3, \ldots). \quad (194)$$

Due to the *continuity* of $F(x)$, (194) implies that $F(x)$ is identically zero, and hence $f(x)$ is also identically zero.

From this theorem it follows that if all the moments of two weight functions coincide, then these two weights are identical.[1] Now it is by no means the case that every arbitrary sequence of numbers is also a sequence of moments of a weight function. as we shall see presently. If this is the case however, then such a sequence will be called a "*moment sequence.*" Still more generally, we shall call a sequence $\{\mu_n\}$ a moment sequence on the interval $[a, b]$ if there exists a function $g(x)$ of *bounded variation* such that

$$\int_a^b x^n dg(x) = \mu_n \qquad (n = 0, 1, 2, \ldots), \quad (195)$$

where the integral is to be understood as a STIELTJES integral. The function $g(x)$ itself will be called an *integral weight* to distinguish it from the *differential weight* $p(x)$ with which we have dealt up until now. The existence of a summable differential weight for which (192) holds immediately implies the existence of a completely continuous integral weight, namely

$$g(x) = \int_a^x p(t)\, dt.$$

If the function $p(x)$ occurring here is positive, then $g(x)$ is *increasing.*

The criterion for determining whether a given sequence is a moment sequence was established by F. HAUSDORFF [1]; the question itself is called the *moment problem.* In the course of this chapter we will set forth the results obtained by HAUSDORFF.

[1] From this it follows that a weight $p(x)$ (up to a constant factor $K > 0$) is uniquely determined by its orthogonal polynomials. An orthogonal system $\{\omega_n(x)\}$ with weight $p(x)$ is also orthogonal with weight $Kp(x)$. Conversely, from

$$\tilde{\omega}_n(x) = x^n + \sigma_1^{(n)} x^{n-1} + \cdots + \sigma_n^{(n)},$$

and putting

$$\mu_0 = \int_a^b p(x)\, dx = 1$$

we obtain the equation

$$\int_a^b p(x)\tilde{\omega}_n(x)\, dx = \mu_n + \sigma_1^{(n)}\mu_{n-1} + \cdots + \sigma_{n-1}^{(n)}\mu_1 + \sigma_n^{(n)} = 0,$$

which determines all the moments.

As is well-known, the value of the STIELTJES integral

$$\int_a^b f(x)\, dg(x) \tag{196}$$

is unchanged when an arbitrary constant is added to $g(x)$. From this fact it is evident that an integral weight is not uniquely determined by its moments. In order to avoid this ambiguity, we shall agree once and for all that for our integral weights

$$g(a) = 0. \tag{197}$$

However, even this does not provide a unique determination of the weight from its moments, since the value of the integral (196) also remains unchanged when we change the value of the function $g(x)$ at any point of the *open interval* (a, b) (this results from the fact that we may omit this point in forming the RIEMANN-STIELTJES sum). To achieve a unique determination of the weight from its moments in spite of this we make the

Definition. A function $g(x)$ of bounded variation on $[a, b]$ shall be called *normal* there if for $a < x < b$

$$g(x) = \frac{g(x-0) + g(x+0)}{2}.$$

Lemma. *For every function $g(x)$ of bounded variation on $[a, b]$ there exists exactly one normal function $g(x)$ of bounded variation which coincides with $g(x)$ at all points of continuity and at the points $x = a$ and $x = b$.*

Proof. We define $\bar{g}(a) = g(a)$, $\bar{g}(b) = g(b)$ and

$$\bar{g}(x) = \frac{g(x-0) + g(x+0)}{2} \qquad (a < x < b).$$

It is obvious that $\bar{g}(x)$ coincides with $g(x)$ at the points of continuity of $g(x)$. Further, $\bar{g}(x)$ is also a function of bounded variation.[2] Since moreover the set of points at which $g(x)$ is continuous is everywhere dense in $[a, b]$

$$\bar{g}(x-0) = g(x-0), \quad \bar{g}(x+0) = g(x+0).$$

[2] Let $\pi(x)$ and $\nu(x)$ be the two monotone increasing components of $g(x)$. If we now introduce two functions $\bar{\pi}(x)$ and $\bar{\nu}(x)$ which coincide with $\pi(x)$ and $\nu(x)$ at the points a and b and in (a, b) are equal to $\frac{\pi(x-0) + \pi(x+0)}{2}$ and $\frac{\nu(x-0) + \nu(x+0)}{2}$ respectively, then $\bar{\pi}(x)$ and $\bar{\nu}(x)$ are monotone increasing. But now $\bar{g}(x) = \bar{\pi}(x) - \bar{\nu}(x)$ and hence $\bar{g}(x)$ is of bounded variation.

Hence, the function $\bar{g}(x)$ is normal. If there were a second normal function satisfying the conditions of the theorem this function would coincide with $\bar{g}(x)$ on an everywhere dense set which would imply that the two functions were identical.

The function $\bar{g}(x)$ just constructed will be called the *kernel* of the function $g(x)$.

Theorem 2. *Two distinct normal integral weights have distinct moment sequences.*

Proof. Suppose two such weights exist whose moment sequences are not distinct; their difference is then again a normal function all the moments of which are zero. Denoting this difference by $g(x)$, we therefore have

$$g(b) - g(a) = \int_a^b dg(x) = 0;$$

hence from (197) $g(b) = 0$.

The rest we obtain through integration by parts:

$$0 = \int_a^b x^n dg(x) = [x^n g(x)]_a^b - n\int_a^b x^{n-1} g(x)\, dx$$

$$(n = 1, 2, 3, \ldots),$$

whence

$$\int_a^b x^{n-1} g(x)\, dx = 0 \qquad (n = 1, 2, 3, \ldots).$$

From Theorem 1 the function $g(x)$ is therefore equivalent to the zero function, and the set of all points at which it vanishes is everywhere dense. Hence $g(x - 0)$ and $g(x + 0)$ are also zero, and since $g(x)$ is a normal function it is identically zero.

A weight and its kernel have the same moments (in forming the RIEMANN-STIELTJES sums it is permissible to choose only points of continuity); the kernel of an integral weight (as distinguished from the weight itself) is therefore uniquely determined by the moments of the weight.

We note further that the solution of the moment problem on the interval [0, 1] is sufficient to obtain the solution on all closed intervals. Let

$$\mu_n = \int_a^b x^n dg(x).$$

We now define

$$h(t) = g[a + t(b - a)].$$

One easily sees that this is a function of bounded variation on [0, 1] whenever $g(x)$ is such a function on $[a, b]$. Since now

$$\mu_n = \int_0^1 [a + t(b - a)]^n \, dh(t),$$

denoting the moments of $h(t)$ by λ_n we obtain

$$\mu_n = \sum_{k=0}^{n} C_n^k a^{n-k} (b - a)^k \lambda_k.$$

Any condition imposed on the numbers μ_n can therefore also be expressed in terms of the λ_n and conversely. We shall therefore restrict ourselves henceforth to the solution of the moment problem on the interval [0, 1].

Finally, let us state without proof two important theorems due to E. HELLY [3] which we shall subsequently need.

I. HELLY'S PRINCIPLE OF CHOICE. *Let a sequence of functions $\{g_n(x)\}$ of bounded variation be given on $[a, b]$. If all the functions of this sequence and their total variations are bounded by a fixed number*

$$|g_n(x)| \leqq K, \qquad \operatorname*{Var}_a^b (g_n) \leqq K,$$

then there exists a sequence of indices $n_1 < n_2 < n_3 < \cdots$ such that the sequence $\{g_{n_k}(x)\}$ converges to a function of bounded variation at all points of $[a, b]$.

II. TRANSITION TO THE LIMIT UNDER THE STIELTJES INTEGRAL SIGN. *Let a sequence of functions $\{g_n(x)\}$ of bounded variation be given on $[a, b]$ which converges to a function of bounded variation $g(x)$ at every point. If the total variations of the functions $g_n(x)$ are bounded by a constant independent of n, then for every continuous function $f(x)$ on $[a, b]$*

$$\lim_{n \to \infty} \int_a^b f(x) dg_n(x) = \int_a^b f(x) dg(x).$$

§2. The Theorems of F. HAUSDORFF

We introduce the following notation: given an arbitrary sequence of real numbers

$$\mu_0, \mu_1, \mu_2, \mu_3, \ldots, \tag{198}$$

we put

$$\Delta^0 \mu_k = \mu_k, \quad \Delta^{n+1} \mu_k = \Delta^n \mu_k - \Delta^n \mu_{k+1}.$$

[3] See, for example, NATANSON [5], p. 222 and p. 233.

Then clearly

$$\Delta^1\mu_k = \mu_k - \mu_{k+1},$$
$$\Delta^2\mu_k = \mu_k - 2\mu_{k+1} + \mu_{k+2}$$

and in general

$$\Delta^n\mu_k = \sum_{i=0}^{n} C_n^i(-1)^i\mu_{k+i}, \tag{199}$$

as is easily proved by induction on n. Comparing (199) with the expansion of $x^k(1-x)^n$ in powers of x, we see that we obtain $\Delta^n\mu_k$ from this expansion by replacing $1, x, \ldots, x^{n+k}$ with the numbers $\mu_0, \mu_1, \ldots, \mu_{n+k}$.

Theorem 1. (F. HAUSDORFF [1]). 1. *If the numbers* (198) *are the moments of an increasing integral weight, then*

$$\Delta^n\mu_k \geqq 0 \tag{200}$$

$$(k = 0, 1, 2, \ldots; \quad n = 0, 1, 2, \ldots).$$

2. *If the numbers* (198) *are the moments of an integral weight of bounded variation, then*

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k}\mu_k| \leqq K \tag{201}$$

$$(n = 0, 1, 2, \ldots),$$

where K does not depend on n.

Proof. In both cases according to our agreement the theorem concerns moments on the interval [0, 1].

Let

$$\mu_k = \int_0^1 x^k dg(x),$$

where $g(x)$ is a function of bounded variation. Then using (199)

$$\int_0^1 x^k(1-x)^n dg(x) = \sum_{i=0}^{n} (-1)^i C_n^i \int_0^1 x^{k+i} dg(x) = \Delta^n\mu_k. \tag{202}$$

If now the function $g(x)$ is increasing, then the integral on the left hand side is non-negative, and this proves the first part of the theorem.

In the general case it follows from (202) that

$$\Delta^{n-k}\mu_k = \int_0^1 x^k(1-x)^{n-k}dg(x).$$

Now setting

$$\varepsilon_k^{(n)} = \operatorname{sign}[\Delta^{n-k}\mu_k],$$

we obtain

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k}\mu_k| = \int_0^1 \left[\sum_{k=0}^{n} \varepsilon_k^{(n)} C_n^k x^k (1-x)^{n-k}\right] dg(x).$$

But for $0 \leqq x \leqq 1$

$$\left|\sum_{k=0}^{n} \varepsilon_k^{(n)} C_n^k x^k (1-x)^{n-k}\right| \leqq 1$$

and hence

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k}\mu_k| \leqq \operatorname*{Var}_0^1 (g),$$

which completes the proof.

It will now be shown that this theorem can also be reversed.

Theorem 2. (F. HAUSDORFF [1]). 1. *If condition* (200) *is satisfied, then the numbers* (198) *constitute the moment sequence of an increasing integral weight.*

2. *If condition* (201) *is satisfied, then the numbers* (198) *constitute the moment sequence of an integral weight of bounded variation.*

We first establish the following: if a sequence of numbers satisfies condition (200), then it also satisfies (201). For if this is the case, then

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k}\mu_k| = \sum_{k=0}^{n} C_n^k \Delta^{n-k}\mu_k.$$

From the remark preceding the first theorem the right-hand side of this equation can be computed by expanding

$$\sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} \tag{203}$$

in powers of x and then replacing x_i by μ_i $(i = 0, 1, 2, \ldots)$.

But the expression (203) is equal to one and therefore

$$\sum_{k=0}^{n} C_n^k \Delta^{n-k}\mu_k = \mu_0.$$

Proof of the theorem. We now introduce step functions $g_n(x)$ which are defined on the interval [0, 1] as follows:

$$g_n(0) = 0,$$

$$g_n(x) = C_n^0 \Delta^n \mu_0 \qquad \text{for } 0 < x \leqq \frac{1}{n},$$

$$g_n(x) = C_n^0 \Delta^n \mu_0 + C_n^1 \Delta^{n-1} \mu_1 \qquad \text{for } \frac{1}{n} < x \leqq \frac{2}{n},$$

. . . . . .

$$g_n(x) = \sum_{k=0}^{n-1} C_n^k \Delta^{n-k} \mu_k \qquad \text{for } \frac{n-1}{n} < x < 1,$$

$$g_n(1) = \sum_{k=0}^{n} C_n^k \Delta^{n-k} \mu_k.$$

If condition (201) is satisfied, then

$$|g_n(x)| \leqq K, \quad \operatorname*{Var}_0^1 (g_n) \leqq K.$$

If condition (200) is satisfied, then $g_n(x)$ is moreover increasing.

From HELLY's principle of choice there now exists a sequence of indices $n_1 < n_2 < n_3 < \cdots$ such that the sequence $\{g_{n_i}(x)\}$ converges to a function $g(x)$ of bounded variation on the entire interval [0, 1]. If all the functions $g_n(x)$ are increasing, then the function $g(x)$ is also increasing.

Then from HELLY's second theorem

$$\int_0^1 x^m dg(x) = \lim_{i \to \infty} \int_0^1 x^m dg_{n_i}(x). \tag{204}$$

But since $g_n(x)$ is a step function with steps at the points $\frac{k}{n}$ where the step heights are $C_n^k \Delta^{n-k} \mu_k$, it follows that

$$\int_0^1 x^m dg_n(x) = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^m C_n^k \Delta^{n-k} \mu_k.$$

The right-hand side of this equation is obtained when the powers x^i occurring in the BERNSTEIN polynomial

$$B_{n,m}(x) = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^m C_n^k x^k (1-x)^{n-k}$$

corresponding to x^m are replaced by the numbers μ_i; i.e. if [4]

$$B_{n,m}(x) = a_0^{(n,m)} + a_1^{(n,m)}x + \cdots + a_m^{(n,m)}x^m ,$$

then it follows that

$$\int_0^1 x^m dg_n(x) = a_0^{(n,m)}\mu_0 + a_1^{(n,m)}\mu_1 + \cdots + a_m^{(n,m)}\mu_m. \tag{205}$$

From BERNSTEIN's theorem

$$\lim_{n\to\infty} B_{n,m}(x) = x^m$$

uniformly on [0, 1].

As we know from Vol. I, it follows from this that the quasinorm of the difference tends to zero:

$$\lim_{n\to\infty}\left[\sum_{i=0}^{m-1}|a_i^{(n,m)}| + |a_m^{(n,m)} - 1|\right] = 0. \tag{206}$$

From (205) and (206) it follows that

$$\lim_{n\to\infty}\int_0^1 x^m dg_n(x) = \mu_m,$$

and according to (204) therefore

$$\int_0^1 x^m dg(x) = \mu_m.$$

The theorem has now been proved, since m was entirely arbitrary.

Theorem 3. (F. HAUSDORFF). *For the existence of the equations*

$$\mu_n = \int_0^1 x^n\varphi(x)\,dx \qquad (n = 0, 1, 2, \ldots), \tag{207}$$

where $\varphi(x)$ is a bounded measurable function, it is necessary and sufficient that for every n

$$(n+1)C_n^k\,|\Delta^{n-k}\mu_k| \leqq K \qquad (k = 0, 1, 2, \ldots, n), \tag{208}$$

where K does not depend on n.

Proof. From (207) and $|\varphi(x)| \leqq K$

$$|\Delta^{n-k}\mu_k| = \left|\int_0^1 x^k(1-x)^{n-k}\varphi(x)\,dx\right| \leqq K\int_0^1 x^k(1-x)^{n-k}\,dx.$$

[4] We recall that the degree of $B_{n,m}$ is m and not n.

But now

$$\int_0^1 x^k(1-x)^{n-k}\,dx = B(k+1, n-k+1)$$

$$= \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{1}{(n+1)C_n^k}, \qquad (209)$$

whence (208) follows.

Now suppose (207) is satisfied. Then (201) is satisfied *a fortiori* and hence

$$\mu_k = \int_0^1 x^k dg(x),$$

where $g(x)$ is a function of bounded variation which we may assume to be normal. Condition (208) then implies that

$$\left| \int_0^1 x^k(1-x)^{n-k} dg(x) \right| \leqq K \int_0^1 x^k(1-x)^{n-k}\,dx.$$

Setting $u(x) = Kx - g(x)$ and $v(x) = Kx + g(x)$, we obtain

$$\int_0^1 x^k(1-x)^{n-k} du(x) \geqq 0, \qquad \int_0^1 x^k(1-x)^{n-k}\,dv(x) \geqq 0.$$

If now λ_k are the moments of $v(x)$, the last equation implies that $\Delta^{n-k}\lambda_k \geqq 0$; therefore [5] the λ_k are the moments of an increasing function. The kernel corresponding to these moments is $v(x)$ (from Theorem 2, § 1), and since the kernel of an increasing function is also increasing, $v(x)$ is increasing. Therefore if $x < y$

$$Kx + g(x) \leqq Ky + g(y)$$

or

$$g(x) - g(y) \leqq K(y - x).$$

With the help of the function $u(x)$ we obtain in the same way the estimate:

$$K(x - y) \leqq g(x) - g(y).$$

[5] We here proceed as in the preceding theorem. Of course the numbers $\Delta^n\mu_k$ appear there and here we are concerned with $\Delta^{n-k}\lambda_k$; however, this clearly makes no difference since $\Delta^m\mu_k = \Delta^{n-k}\mu_k$ if $n = m + k$.

The function $g(x)$ therefore satisfies a LIPSCHITZ condition with coefficient K. But then

$$g(x) = \int_0^x \varphi(t)\, dt,$$

where $|\varphi(x)| \leqq K$. From this we obtain (207).

Theorem 4. (F. HAUSDORFF). *If $\varphi(x) \in L^2$ the necessary and sufficient condition for the existence of equations (207) is that*

$$(n+1) \sum_{k=0}^{n} [C_n^k \Delta^{n-k} \mu_k]^2 \leqq K \qquad (n = 0, 1, 2, \ldots) \quad (210)$$

where K does not depend on n.

Proof. Suppose (207) holds and $\varphi(x) \in L^2$, then

$$\Delta^{n-k} \mu_k = \int_0^1 x^k (1-x)^{n-k} \varphi(x)\, dx.$$

The SCHWARZ inequality yields

$$(\Delta^{n-k} \mu_k)^2 \leqq \left\{ \int_0^1 x^k (1-x)^{n-k}\, dx \right\} \left\{ \int_0^1 x^k (1-x)^{n-k} \varphi^2(x)\, dx \right\},$$

whence using (209) we obtain the estimate:

$$(n+1) C_n^k (\Delta^{n-k} \mu_k)^2 \leqq \int_0^1 x^k (1-x)^{n-k} \varphi^2(x)\, dx.$$

Therefore

$$(n+1) \sum_{k=0}^{n} [C_n^k \Delta^{n-k} \mu_k]^2 \leqq \int_0^1 \left[\sum_{k=0}^{n} C_n^k x^k (1-x)^{n-k} \right] \varphi^2(x)\, dx = \int_0^1 \varphi^2(x)\, dx.$$

This proves the necessity of condition (210). That the condition is also sufficient will be proved in the next section.

§3. Linear Functionals on C and L^2

To every continuous function $f(x)$ on $[a, b]$ let there correspond a real number $\Phi(f)$ such that

$$\Phi(f_1 + f_2) = \Phi(f_1) + \Phi(f_2), \quad (211)$$

$$|\Phi(f)| \leqq K \max |f(x)| \quad (212)$$

where K is independent of f. $\Phi(f)$ is then called a *linear functional* on $C([a, b])$.

An example is furnished by every STIELTJES integral

$$\Phi(f) = \int_a^b f(x)dg(x) \tag{213}$$

in which $g(x)$ is a function of bounded variation. To satisfy (212) one need only set $K = \mathrm{Var}\ (g)$.

If a linear functional is non-negative for non-negative functions $f(x)$ it is called a positive linear functional. The integral (213) in which the integrating function is monotone increasing is an example.

It will turn out that there exist no linear functionals other than those of the form (213). This we shall prove in the present section.

Lemma. *If $\Phi(f)$ is a linear functional, then*

$$\Phi(kf) = k\Phi(f). \tag{214}$$

Proof. From (212) it follows that $\Phi(0) = 0$ and hence (214) is verified for the case $k = 0$. For any positive integer k the required relation is easily established by induction using (211). Therefore, it clearly holds for $k = \frac{1}{n}$ also, where n is a positive integer. From this it further follows that (214) is valid for all positive rational numbers. If k is a negative rational number, then

$$\Phi(kf) + \Phi(-kf) = \Phi(0) = 0,$$

so that (214) holds for all rational numbers k.

Finally, if k is an irrational and r a rational number then

$$|\Phi(kf) - k\Phi(f)| \leqq |\Phi[(k-r)f]| + |k-r|\,|\Phi(f)|,$$

and hence from (212)

$$|\Phi(kf) - k\Phi(f)| \leqq 2K|k-r| \max |f(x)|,$$

whence we obtain (214), since the right-hand side of the last equation can be made arbitrarily small.

Theorem 1. (F. RIESZ [2]). *Every linear functional $\Phi(f)$ on $C([a, b])$ can be represented in the form* (213) *where $g(x)$ is a function of bounded variation.*

Proof. We note first that it suffices to prove the theorem for the interval $[0, 1]$, since the general case can be reduced to this special case by a linear transformation. But now for this special case the RIESZ theorem can be easily obtained from the HAUSDORFF theorem.

Let $\Phi(f)$ be any linear functional on $C([0, 1])$. Let us put

$$\Phi(x^n) = \mu_n \qquad (n = 0, 1, 2, \ldots), \quad (215)$$

then

$$\Delta^n \mu_k = \sum_{i=0}^{n} (-1)^i C_n^i \mu_{k+i} = \Phi\left[\sum_{i=0}^{n} (-1)^i C_n^i x^{k+i}\right] = \Phi[x^k(1 - x)^n]. \quad (216)$$

From this it follows that

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k} \mu_k| = \Phi\left[\sum_{k=0}^{n} \varepsilon_k^{(n)} C_n^k x^k (1 - x)^{n-k}\right],$$

where $\varepsilon_k^{(n)} = \text{sign } (\Delta^{n-k} \mu_k)$.

For $0 \leqq x \leqq 1$

$$\left| \sum_{k=0}^{n} \varepsilon_k^{(n)} C_n^k x^k (1 - x)^{n-k} \right| \leqq 1;$$

condition (212) then gives

$$\sum_{k=0}^{n} C_n^k |\Delta^{n-k} \mu_k| \leqq K.$$

The numbers μ_k therefore satisfy condition (201), so that there exists a function $g(x)$ of bounded variation of which they are the moments. Therefore

$$\Phi(x^n) = \int_0^1 x^n dg(x) \qquad (n = 0, 1, 2, \ldots). \quad (217)$$

From this it follows first of all that equation (213) holds when $f(x)$ is any arbitrary polynomial; but since every continuous function is the limit of a uniformly convergent sequence of polynomials, the theorem is herewith proved in general.

Still simpler is the proof of

Theorem 2. (F. Riesz). *Every positive linear functional on $C([a, b])$ is represented by*

$$\Phi(f) = \int_a^b f(x) dg(x)$$

where $g(x)$ is an increasing function.

Proof. We again restrict ourselves to the interval [0, 1] and again introduce the numbers (215). From (216) and the inequality $x^k(1 - x)^{n-k} \geqq 0$ which is valid for $0 \leqq x \leqq 1$ we obtain $\Delta^n \mu_k \geqq 0$. Therefore, an increasing function

$g(x)$ can be found which satisfies conditions (217); the remainder of the proof goes as before.

In addition to linear functions on C we shall also need linear functionals on L^2.

If to every square-summable function $f(x)$ on $[a, b]$ there is made to correspond a number $\Phi(f)$ such that

$$\Phi(f_1 + f_2) = \Phi(f_1) + \Phi(f_2), \qquad |\Phi(f)| \leqq K \sqrt{\int_a^b f^2(x)\, dx},$$

then $\Phi(f)$ is called a linear functional on L^2 (the space L^2 depends on the interval of definition). Relation (214) holds here just as for the space C.

An example of a linear functional on L^2 is given by

$$\Phi(f) = \int_a^b f(x)\varphi(x)\, dx,$$

where $\varphi(x)$ itself belongs to L^2. It will again be shown that this is the general form of such functionals, for we have

Theorem 3. (M. FRÉCHET [1]). *Every linear functional $\Phi(f)$ on L^2 can be represented in the form*

$$\Phi(f) = \int_a^b f(x)\varphi(x)\, dx, \tag{218}$$

where $\varphi(x) \in L^2$.

Proof. We select an arbitrary complete orthonormal system $\{\omega_k(x)\}$ and put

$$\Phi(\omega_k) = A_k.$$

Then

$$\sum_{k=1}^{n} A_k^2 = \Phi\left(\sum_{k=1}^{n} A_k\omega_k\right) \leqq K \sqrt{\int_a^b \left[\sum_{k=1}^{n} A_k\omega_k(x)\right]^2 dx} = K\sqrt{\sum_{k=1}^{n} A_k^2}$$

and hence

$$\sum_{k=1}^{\infty} A_k^2 \leqq K^2.$$

Now if this is the case there exists a square-summable function $\varphi(x)$ whose FOURIER coefficients with respect to the system $\{\omega_k(x)\}$ are just the numbers

A_k. Moreover, this function satisfies condition (218). For if $f(x)$ is an arbitrary function of L^2 and if $\{c_k\}$ are its FOURIER coefficients, then

$$\lim_{n\to\infty} \int_a^b \left[f(x) - \sum_{k=1}^{n} c_k \omega_k(x) \right]^2 dx = 0.$$

But

$$\left| \Phi(f) - \Phi\left(\sum_{k=1}^{n} c_k \omega_k \right) \right| \leqq K \left\| f - \sum_{k=1}^{n} c_k \omega_k \right\|,$$

and therefore

$$\Phi(f) = \lim_{n\to\infty} \Phi\left(\sum_{k=1}^{n} c_k \omega_k \right) = \lim_{n\to\infty} \sum_{k=1}^{n} A_k c_k = \sum_{k=1}^{\infty} A_k c_k;$$

from the generalized PARSEVAL equality the last expression is equal to

$$\int_a^b f(x)\varphi(x)\, dx.$$

This theorem now makes it possible for us to complete the proof of the fourth HAUSDORFF theorem in § 2. Let $\{\mu_n\}$ be a sequence of numbers satisfying condition (210). The CAUCHY inequality then provides the estimate

$$\sum_{k=0}^{n} C_n^k \left| \Delta^{n-k} \mu_k \right| \leqq \sqrt{n+1} \sqrt{\sum_{k=0}^{n} \left[C_n^k \Delta^{n-k} \mu_k \right]^2} \leqq \sqrt{K},$$

and the second HAUSDORFF theorem supplies a function $g(x)$ of bounded variation for which

$$\int_0^1 x^n dg(x) = \mu_n \qquad (n = 0, 1, 2, \ldots).$$

We now take any continuous function $f(x)$ and form its BERNSTEIN polynomial

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}.$$

Then

$$\int_0^1 B_n(x) dg(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k \Delta^{n-k} \mu_k$$

and from the CAUCHY inequality

$$\left[\int_0^1 B_n(x) dg(x) \right]^2 \leqq \left\{ \sum_{k=0}^{n} f^2\left(\frac{k}{n}\right) \right\} \left\{ \sum_{k=0}^{n} \left[C_n^k \Delta^{n-k} \mu_k \right]^2 \right\}.$$

From this and (210) we obtain

$$\left[\int_0^1 B_n(x)dg(x)\right]^2 \leqq K \sum_{k=0}^{n} f^2\left(\frac{k}{n}\right)\frac{1}{n}$$

and by passing to the limit $n \to \infty$:

$$\left|\int_0^1 f(x)dg(x)\right| \leqq \sqrt{K}\sqrt{\int_0^1 f^2(x)\,dx}. \tag{219}$$

Therefore

$$\int_0^1 f(x)dg(x) = \Phi(f)$$

is a linear functional on the space C; we now extend it to the entire space L^2. For this purpose we select an arbitrary function $f(x)$ from L^2; there exists a sequence of continuous functions $f_n(x)$ converging in mean to $f(x)$.

This sequence also converges in itself; but since

$$|\Phi(f_n) - \Phi(f_m)| \leqq \sqrt{K}\sqrt{\int_0^1 (f_n - f_m)^2\,dx},$$

it follows that

$$\lim_{n\to\infty} \Phi(f_n)$$

exists as a finite limit. It is easily seen that this limit does not depend on the choice of the sequence of continuous functions which were used in establishing it.[6] Further, the functional $\Phi(f)$ is obviously linear[7]; using FRÉCHET's theorem we represent it in the form (218) with $\varphi(x) \in L^2$. From this it follows that

$$\int_0^1 x^n\varphi(x)\,dx = \Phi(x^n) = \int_0^1 x^n dg(x) = \mu_n,$$

wherewith the proof of the HAUSDORFF theorem is completed.

[6] Indeed, if $\{f_n'(x)\}$ is another sequence of continuous functions converging to $f(x)$ in mean, then the difference $f_n(x) - f_n'(x)$ converges in mean to zero, so that (from 219) $\lim [\Phi(f_n) - \Phi(f_n')] = 0$.

[7] If $f'(x)$ and $f''(x)$ are two functions of L^2 and $\{f_n'(x)\}$ and $\{f_n''(x)\}$ respectively are two sequences of continuous functions converging to them in mean, then the sequence $\{f_n' + f_n''\}$ converges in mean to $f'(x) + f''(x)$, so that $\Phi(f' + f'') = \Phi(f') + \Phi(f'')$. The inequality $|\Phi(f)| \leqq \sqrt{K}\,\|f\|$ is obtained by taking the limit in (219).

§4. Positive Definite Sequences

The RIESZ theorem concerning linear functionals on the space C furnishes us with a characteristic property of the moment sequences of an increasing integral weight.

We shall call a sequence $\{\mu_n\}$ $(n = 0, 1, 2, \ldots)$ of real numbers *positive definite* on the interval [8] $<a, b>$ if for every polynomial

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n,$$

which is non-negative and not identically zero on $<a, b>$ the inequality

$$\Phi(P) = a_0\mu_0 + a_1\mu + \cdots + a_n\mu_n \geqq 0$$

holds.

If for every such polynomial $\Phi(P) > 0$ we shall call the sequence $\{\mu_n\}$ *strictly* positive definite on $<a, b>$.

Theorem 1. *A sequence of numbers $\{\mu_n\}$ is the moment sequence of an increasing integral weight on $[a, b]$*

$$\mu_n = \int_a^b x^n dg(x) \qquad (n = 0, 1, 2, \ldots),$$

if and only if it is positive definite on $[a, b]$.

Proof. If we have an increasing integral weight $g(x)$, then for all polynomials which are non-negative on $[a, b]$

$$\Phi(P) = \int_a^b P(x) dg(x),$$

so that the sequence $\{\mu_n\}$ is positive definite on $[a, b]$.

Now suppose $\{\mu_n\}$ is positive definite on $[a, b]$. Let us choose an arbitrary polynomial

$$P(x) = \sum_{k=0}^{n} a_k x^k$$

and put

$$M = \max_{a \leqq x \leqq b} |P(x)|,$$

so that both of the polynomials

$$M + P(x), \quad M - P(x)$$

[8] This can be a closed, half-open, or open interval.

are non-negative on $[a, b]$; hence

$$(M + a_0)\mu_0 + a_1\mu_1 + \cdots + a_n\mu_n \geqq 0,$$
$$(M - a_0)\mu_0 - a_1\mu_1 - \cdots - a_n\mu_n \geqq 0.$$

From this it follows that

$$|\Phi(P)| \leqq \mu_0 \max |P(x)|. \tag{220}$$

Now let $f(x)$ be any continuous function on $[a, b]$; the WEIERSTRASS theorem then asserts the existence of a sequence of polynomials $\{P_n(x)\}$ converging uniformly to $f(x)$. Therefore, for every given $\varepsilon > 0$ an N can be found such that for all $n > N$

$$|P_n(x) - f(x)| < \frac{\varepsilon}{2};$$

for $n > N$ and $m > N$ therefore

$$|P_n(x) - P_m(x)| < \varepsilon,$$

and from (220)

$$|\Phi(P_n) - \Phi(P_m)| < \mu_0 \varepsilon.$$

Hence, the sequence $\{\Phi(P_n)\}$ converges in itself and has a finite limit which we denote by $\Phi(f)$. It is easily seen that $\Phi(f)$ depends only on the function $f(x)$ and the sequence $\{\mu_n\}$, but not on the sequence of polynomials $\{P_n(x)\}$ selected. Indeed, if $\{Q_n(x)\}$ is another sequence of polynomials converging uniformly to $f(x)$, then

$$|\Phi(P_n) - \Phi(Q_n)| \leqq \mu_0 \max |P_n(x) - Q_n(x)|,$$

where the right-hand side of this inequality tends to zero.

We have thus defined a functional $\Phi(f)$ on the set of all continuous functions on $[a, b]$, and we shall see immediately that it is a positive linear functional. If $f(x) = f_1(x) + f_2(x)$ we have two uniformly converging sequences:

$$P_n^{(1)}(x) \rightrightarrows f_1(x), \qquad P_n^{(2)}(x) \rightrightarrows f_2(x),$$

hence $P_n^{(1)}(x) + P_n^{(2)}(x) \rightrightarrows f(x)$, and therefore

$$\Phi(f) = \lim \Phi[P_n^{(1)} + P_n^{(2)}] = \Phi(f_1) + \Phi(f_2).$$

Now if $f(x) \geqq 0$ all the polynomials $P_n(x)$ may be chosen to be non-negative,[9] and hence it follows that $\Phi(f) \geqq 0$.

[9] Let $\{Q_n(x)\}$ be any sequence of polynomials converging uniformly to $f(x)$. We set $\rho_n = \max |Q_n(x) - f(x)|$, where clearly $\rho_n \to 0$. The polynomials

$$P_n(x) = Q_n(x) + \rho_n$$

are then non-negative and converge uniformly to $f(x)$.

If finally $f(x)$ is any continuous function on $[a, b]$ and $M = \max |f(x)|$, then both functions $M \pm f(x)$ are non-negative, whence

$$M\mu_0 \pm \Phi(f) \geqq 0$$

and therefore

$$|\Phi(f)| \leqq \mu_0 M.$$

$\Phi(f)$ is hence a positive linear functional, and from the RIESZ theorem

$$\Phi(f) = \int_a^b f(x)dg(x).$$

The function $g(x)$ occurring here is increasing, and it remains only to remark that $\Phi(x^n) = \mu_n$.

Before beginning the study of strictly positive definite sequences we first make the

Definition. Let $g(x)$ be an increasing function on $[a, b]$. We shall call a point x_0 $(a < x_0 < b)$ a *point of increase* of the function $g(x)$ if for every pair of points y, z satisfying the inequalities $a \leqq y < x_0 < z \leqq b$, $g(y) < g(z)$. If $g(a) < g(x)$ for all $x \in (a, b]$ the point a is called a point of increase; if $g(x) < g(b)$ for all $x \in [a, b)$ the point b is called a point of increase.

Lemma 1. *A nonconstant increasing function $g(x)$ has at least one point of increase.*

Proof. By hypothesis $g(b) > g(a)$. If we set $c = \dfrac{a+b}{2}$, then at least one of the two differences $g(c) - g(a)$ and $g(b) - g(c)$ is positive. We now denote by $[a_1, b_1]$ that one of the two intervals $[a, c]$ and $[c, b]$ for which $g(a_1) < g(b_1)$. By continuing this procedure we obtain a sequence of nested intervals $\{[a_n, b_n]\}$ such that $g(a_n) < g(b_n)$ for all n. The common point x_0 of these closed intervals is then a point of increase of $g(x)$.

Corollary. *If a function $g(x)$ which is increasing on $[a, b]$ has only a finite number of points of increase, then it is a step function which is constant between points of increase.*

Lemma 2. *If $g(x)$ is an increasing function with infinitely many points of increase, then for every non-negative polynomial $P(x)$ which is not identically zero the inequality*

$$\int_a^b P(x)dg(x) > 0.$$

holds.

If on the other hand $g(x)$ has only a finite number of points of increase, then there exists a non-negative polynomial $P(x)$, not identically zero, for which

$$\int_a^b P(x)dg(x) = 0.$$

Proof. In the first case a point of increase x_0 of $g(x)$ can be found which is not a root of $P(x)$. Now let $h > 0$ be so small that $[x_0 - h, x_0 + h]$ is contained in $[a, b]$ and no root of $P(x)$ lies in $[x_0 - h, x_0 + h]$. Denoting the minimum value of $P(x)$ on $[x_0 - h, x_0 + h]$ by m, we have

$$\int_a^b P(x)dg(x) \geqq \int_{x_0-h}^{x_0+h} P(x)dg(x) \geqq m[g(x_0 + h) - g(x_0 - h)] > 0\ .^{10}$$

In the second case let $\xi_1, \xi_2, \ldots, \xi_s$ be the points of increase of $g(x)$. If we put

$$P(x) = (x - \xi_1)^2(x - \xi_2)^2 \ldots (x - \xi_s)^2,$$

then $P(x) \geqq 0$ and

$$\int_a^b P(x)dg(x) = \sum_{k=1}^{s} P(\xi_k)[g(\xi_k + 0) - g(\xi_k - 0)] = 0.$$

We have now obviously established

Theorem 2. *A sequence of numbers $\{\mu_n\}$ is the moment sequence of an increasing function $g(x)$ with infinitely many points of increase in $[a, b]$, i.e.*

$$\int_a^b x^n dg(x) = \mu_n \qquad (n = 0, 1, 2, \ldots),$$

if and only if it is strictly positive definite on $[a, b]$.

We must note the following: the properties of the moment sequences established here are less useful than those found by HAUSDORFF, since we are not in possession of a criterion to determine whether a given sequence is positive definite or not. We will later see that such a criterion exists for the infinite interval $(-\infty, +\infty)$.[11]

[10] This shows still more: namely that the inequality $\int_a^b P\,dg > 0$ is still satisfied when the number of points of increase of $g(x)$ is greater than the degree of $P(x)$. We will later make use of this fact.

[11] One can by the way consider the HAUSDORFF condition $\Delta^n\mu_k \geqq 0$ to be a useful criterion to determine if the sequence $\{\mu_k\}$ is positive definite on [0, 1].

The reader who wishes to go into the questions of this chapter more deeply will find the monograph by N. I. AKHIEZER and M. G. KREIN [2] useful [12]; it is very rich in content, but somewhat hard to read.

[12] See also L. V. KANTOROVICH [2].

CHAPTER VIII

INFINITE INTERVALS

§1. Introduction

All investigations thus far have involved a finite interval $[a, b]$. We shall now consider infinite intervals; here everything will be different from the beginning, since the WEIERSTRASS theorem no longer holds and all results based on it are lost.

As heretofore, we consider the classes $L_{p(x)}$ and $L^2_{p(x)}$ where $p(x)$ is a non-negative measurable weight function which vanishes on a set of measure zero at most.[1] In addition, we assume that for the weight $p(x)$ all moments

$$\mu_n = \int_a^b x^n p(x)\, dx \qquad (n = 0, 1, 2, \ldots)$$

exist. (For finite intervals the existence of all the moments follows from the summability of $p(x)$. This is no longer the case. For example, the weight $p(x) = \dfrac{1}{1 + x^2}$ is summable on $(-\infty, +\infty)$, but μ_n does not exist for $n > 0$.)

We say that the function $f(x)$ belongs to $L_{p(x)}$ if it is measurable and if

$$\int_a^b |f(x)|\, p(x)\, dx < +\infty,$$

where the integral is to be understood as an improper integral, e.g. for $a = 0$, $b = +\infty$ as

$$\lim_{A \to +\infty} \int_0^A |f(x)|\, p(x)\, dx.$$

[1] A function defined on an infinite interval is called measurable if it is measurable on every finite subinterval of the given interval. An unbounded set is called measurable if its intersection with every finite interval is measurable. It has measure zero if its intersection with all finite intervals are sets of measure zero.

If $f(x)$ is measurable and $f^2(x) \in L_{p(x)}$ we say that $f(x)$ belongs to $L^2_{p(x)}$. As before [2] $L^2_{p(x)} \subset L_{p(x)}$, since $|f(x)| \leqq \frac{1}{2}[1 + f^2(x)]$.

The CAUCHY and SCHWARZ inequalities and all geometrical relations based on them hold as before. Also, all bounded measurable functions belong to $L_{p(x)}$ as formerly, but all continuous functions no longer do.

If for example $a = 0$, $b = +\infty$, $p(x) = e^{-x}$ (the "LAGUERRE" weight), then $f(x) = e^x$ does not belong to $L_{e^{-x}}$ and hence does not belong to $L^2_{e^{-x}}$.

There is nevertheless the

Theorem. *The class C of all bounded continuous functions is everywhere dense in $L^2_{p(x)}$.*

Proof. We prove the theorem for $a = 0$, $b = +\infty$ by way of example. Suppose $f(x) \in L^2_{p(x)}$ and $\varepsilon > 0$. Then for sufficiently large values of A

$$\int_A^{+\infty} p(x)f^2(x)\,dx < \frac{\varepsilon^2}{2}.$$

Since the theorem has already been proved for finite intervals, it is possible to find a continuous function $\psi_0(x)$ defined on $[0, A]$ for which

$$\int_0^A p(x)[f(x) - \psi_0(x)]^2\,dx < \frac{\varepsilon^2}{8}.$$

Now let $\delta < A$ be a positive number which we shall subsequently specify more precisely. We construct a function $\psi(x)$ by defining $\psi(A) = 0$, $\psi(x) = \psi_0(x)$ for $0 \leqq x \leqq A - \delta$, and $\psi(x)$ linear on the interval $[A - \delta, A]$. If $M = \max |\psi_0(x)|$, then

$$\int_0^A p(x)[\psi_0(x) - \psi(x)]^2\,dx \leqq 4M^2 \int_{A-\delta}^A p(x)\,dx.$$

We can now choose δ so small that the right-hand side of the last equation becomes less than $\frac{\varepsilon^2}{8}$. Then

$$\int_0^A p(x)[f(x) - \psi(x)]^2\,dx < \frac{\varepsilon^2}{2}.$$

[2] If we do not assume that $p(x)$ is summable, then this is not true. As an example, for $a = 1$, $b = \infty$, $p(x) = 1$ the function $\frac{1}{x}$ belongs to L^2 but not to L.

If we define $\varphi(x) = \begin{cases} \psi(x) & \text{for } 0 \leqq x \leqq A, \\ 0 & \text{for } x > A, \end{cases}$

then $\varphi(x)$ is a bounded continuous function for which

$$\|f - \varphi\| < \varepsilon,$$

wherewith the theorem is proved.

Remark. From the proof of the theorem it is also clear that the class of all those continuous functions which are zero from some point on is everywhere dense in $L^2_{p(x)}$.

The class of all polynomials however need not be dense in $L^2_{p(x)}$. This is demonstrated by an

Example.[3] Suppose $a = 0$, $b = +\infty$, $p(x) = x^{-\ln x}$. The function $p(x)$ is continuous and bounded, since

$$\ln p(x) = -\ln^2 x;$$

for $x \to 0$ and $x \to +\infty$ therefore $p(x) \to 0$.

Further, $p(x)$ possesses all its moments, since $p(x)$ is summable on every finite interval $[0, A]$.

Now for $x > e^{n+2}$

$$x^n p(x) < \frac{1}{x^2},$$

so that

$$\int_A^{+\infty} x^n p(x)\, dx \qquad (A = e^{n+2})$$

exists. Hence, $p(x)$ is a weight function of the type described above.

Now suppose

$$f(x) = \sin(2\pi \ln x).$$

This function is bounded and therefore belongs to $L^2_{p(x)}$. We shall now show that with the weight $p(x)$ it is orthogonal to all functions x^n $(n = 0, 1, 2, \ldots)$. With the substitution

$$x = e^{t + \frac{n+1}{2}}$$

we obtain

$$\int_0^{+\infty} x^n f(x) p(x)\, dx = (-1)^{n+1} e^{\left(\frac{n+1}{2}\right)^2} \int_{-\infty}^{+\infty} e^{-t^2} \sin 2\pi t\, dt,$$

and the last integral is obviously zero, since the integrand is an *odd* function.

[3] STIELTJES [2].

Therefore, for every polynomial $P(x)$ the equation

$$\int_0^{+\infty} f(x)P(x)p(x)\,dx = 0$$

holds.

Therefore

$$\int_0^{+\infty} p(x)f^2(x)\,dx = \int_0^{+\infty} p(x)f(x)[f(x) - P(x)]\,dx$$

and hence from the SCHWARZ inequality

$$\left\{\int_0^{+\infty} p(x)f^2(x)\,dx\right\}^2 \leqq \left\{\int_0^{+\infty} p(x)f^2(x)\,dx\right\}\left\{\int_0^{+\infty} p(x)[f(x) - P(x)]^2\,dx\right\},$$

whence

$$\|f - P\| \geqq \|f\|,$$

The quadratic approximation of any polynomial to the function $f(x)$ is therefore not smaller than $\|f\| > 0$.

In spite of this negative result there exist weights $p(x)$ in the space $L^2_{p(x)}$ of which the class of all polynomials is dense. We will later see that the LAGUERRE weight e^{-x} in $[0, +\infty)$ and the HERMITE weight e^{-x^2} in $(-\infty, +\infty)$ have this property. Without getting lost in details we can say that the weight function must decrease sufficiently fast as $x \to \infty$.

We note further that $L^2_{p(x)}$ is also complete for infinite intervals: every sequence which converges in itself has a limit in $L^2_{p(x)}$. This is proved just as in the case of finite intervals.

The theory of orthogonal systems including the RIESZ-FISCHER theorem can also be constructed in the same manner. In particular, for orthogonal systems completeness and closedness are equivalent properties, as are completeness and the existence of fundamental sequences for arbitrary linearly independent systems.

The extremal property of the FOURIER coefficients (the TOEPLER theorem), the BESSEL inequality, the SCHMIDT orthogonalization process, and etc. hold as before.

In particular, it is possible as formerly to obtain orthogonal systems of polynomials by orthogonalizing the system of powers x^n (the existence of all the moments of the weight function is hereby essential). We will however not concern ourselves with the general properties of such polynomial systems, but will rather in the next sections restrict ourselves to two special cases: the LAGUERRE and the HERMITE polynomials.

§2. The LAGUERRE Polynomials

The polynomials forming an orthogonal system on the interval $[0, +\infty]$ with weight e^{-x} are called LAGUERRE *polynomials.*[4] They are represented (up to an arbitrary constant factor) by the formula

$$L_n(x) = e^x \frac{d^n}{dx^n}(x^n e^{-x}),$$

which is analogous to the RODRIGUES formula.

To prove this assertion we set

$$U_n(x) = x^n e^{-x}.$$

The LEIBNIZ formula gives

$$U_n^{(m)}(x) = \sum_{k=0}^{m} (-1)^{m-k} C_m^k n(n-1)\dots(n-k+1)x^{n-k}e^{-x}. \quad (221)$$

From this it follows that $L_n(x)$ is indeed a polynomial of degree n which admits the explicit representation

$$L_n(x) = \sum_{k=0}^{n} (-1)^{n-k} C_n^k n(n-1)\dots(n-k+1)x^{n-k}. \quad (222)$$

In particular, the leading coefficient of $L_n(x)$ is $(-1)^n$ so that

$$L_n(x) = (-1)^n e^x \frac{d^n(x^n e^{-x})}{dx^n}.$$

For $m = 0, 1, \dots, n-1$ equation (221) implies that

$$U_n^{(m)}(0) = U_n^{(m)}(+\infty) = 0.$$

Thus if $V(x)$ is a polynomial of H_{n-1}, t hen

$$\int_0^{+\infty} e^{-x} V(x) L_n(x)\,dx = \int_0^{+\infty} U_n^{(n)} V\,dx$$

$$= [U_n^{(n-1)}V - \dots + (-1)^{n-1} U_n V^{(n-1)}]_0^{+\infty}$$

$$+ (-1)^n \int_0^{+\infty} U_n V^{(n)}\,dx = 0.$$

Hence, the polynomials $L_n(x)$ do indeed form an orthogonal system with weight e^{-x}.

[4] LAGUERRE [1].

If on the other hand we put $V(x) = L_n(x)$ in the last equation we obtain

$$\int_0^{+\infty} e^{-x}[L_n(x)]^2\,dx = (-1)^n \int_0^{+\infty} U_n V^{(n)}\,dx,$$

and since $V^{(n)}(x) = (-1)^n n!$

$$\int_0^{+\infty} e^{-x} L_n^2(x)\,dx = n! \int_0^{+\infty} x^n e^{-x}\,dx = (n!)^2, \tag{223}$$

and hence

$$\tilde{L}_n(x) = \frac{(-1)^n}{n!}\, e^x \frac{d^n(x^n e^{-x})}{dx^n}.$$

In the recursion formula

$$\tilde{L}_{n+2}(x) = (x - \alpha_{n+2})\tilde{L}_{n+1}(x) - \lambda_{n+1}\tilde{L}_n(x) \tag{224}$$

according to the general theory

$$\lambda_{n+1} = \frac{\int_0^{+\infty} e^{-x}\tilde{L}_{n+1}^2(x)\,dx}{\int_0^{+\infty} e^{-x}\tilde{L}_n^2(x)\,dx},$$

whence using (223) $\lambda_{n+1} = (n+1)^2$. The number α_{n+2} can be most easily determined by comparing the coefficients of x^{n+1} in (224). With the help of (222) we find $\alpha_{n+2} = 2n + 3$, so that formula (224) now assumes the form

$$\tilde{L}_{n+2}(x) = [x - (2n+3)]\tilde{L}_{n+1}(x) - (n+1)^2\tilde{L}_n(x) \tag{225}$$

Since $\tilde{L}_0(x) = 1$ and $\tilde{L}_1(x) = x - 1$, we see from this that

$$\begin{aligned} \tilde{L}_2(x) &= x^2 - 4x + 2, \\ \tilde{L}_3(x) &= x^3 - 9x^2 + 18x - 6, \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

Now let us consider the function

$$L(t, x) = \frac{1}{1+t} e^{\frac{xt}{1+t}}, \tag{226}$$

which, as is easily seen, for $|t| < 1$ can be expanded in a series of powers of t:

$$L(t, x) = A_0(x) + A_1(x)t + A_2(x)t^2 + \cdots;$$

we shall show that $A_n(x) = \tilde{L}_n(x)$. Obviously $A_0(x) = 1$. Moreover

$$\frac{\partial L(t, x)}{\partial t} = \left[\frac{x}{(1+t)^2} - \frac{1}{1+t}\right] L(t, x), \tag{227}$$

whence it follows that $A_1(x) = x - 1 = \tilde{L}_1(x)$. Writing (227) in the form

$$(1+t)^2 L_t'(t, x) = (x - t - 1)L(t, x)$$

and comparing coefficients of t^{n+1}, we find

$$(n+2)A_{n+2}(x) = [x - (2n+3)]A_{n+1}(x) - (n+1)A_n(x)$$

$$(n = 0, 1, 2, \ldots).$$

If we now multiply by $(n+1)!$, then comparison with (225) gives immediately

$$n!A_n(x) = \tilde{L}_n(x),$$

wherewith our assertion is proved. The function (226) is hence the *generating function* for the LAGUERRE *polynomials.*

It can be easily shown that the LAGUERRE polynomials $L_n(x)$ satisfy the differential equation

$$xy'' + (1 - x)y' + ny = 0. \tag{228}$$

Indeed, if we put

$$u = x^n e^{-x},$$

then $xu' = (n - x)u$. Differentiating this equation $n + 1$ times, the LEIBNIZ formula gives

$$xu^{(n+2)} + (x+1)u^{(n+1)} + (n+1)u^{(n)} = 0. \tag{229}$$

But $u^{(n)} = e^{-x}L_n(x) = e^{-x}y$. From this and (229) one obtains (228).

§3. The Generalized LAGUERRE Polynomials

In many investigations a generalization of the polynomials just considered plays a role: the *generalized* LAGUERRE *polynomials* $L_n^{(\alpha)}(x)$ which for $\alpha > -1$ form an orthogonal system on the interval $[0, +\infty)$ with weight $x^\alpha e^{-x}$. We list briefly their simplest properties; these can be derived in the same manner as for the ordinary LAGUERRE polynomials.

1. The polynomials $L_n^{(\alpha)}(x)$ are represented by the formula

$$L_n^{(\alpha)}(x) = K_n x^{-\alpha} e^x \frac{d^n}{dx^n}(x^{\alpha+n}e^{-x})$$

2. The polynomials $\tilde{L}_n^{(\alpha)}(x)$ with leading coefficient equal to one are obtained for $K_n = (-1)^n$, and the normalized polynomials $\hat{L}_n^{(\alpha)}(x)$ are obtained for

$$K_n = \frac{(-1)^n}{\sqrt{n!\Gamma(\alpha + n + 1)}}.$$

3. The recursion formula for the $\tilde{L}_n^{(\alpha)}(x)$ reads

$$\tilde{L}_{n+2}^{(\alpha)}(x) = (x - \alpha - 2n - 3)\tilde{L}_{n+1}^{(\alpha)}(x) - (n + 1)(n + \alpha + 1)\tilde{L}_n^{(\alpha)}(x).$$

4. For these polynomials

$$L^{(\alpha)}(t, x) = \frac{1}{(1 + t)^{\alpha+1}} e^{\frac{xt}{1+t}} = \sum_{n=0}^{\infty} \frac{\tilde{L}_n^{(\alpha)}(x)}{n!} t^n.$$

serves as generating function.
We now prove the

Theorem. (V. A. STEKLOV [4]). *The set of all polynomials is everywhere dense in the space $L^2_{p(x)}$ of the* LAGUERRE *weight function*

$$p(x) = x^\alpha e^{-x} \qquad (0 \leqq x < +\infty, \alpha > -1).$$

Proof. First of all, from the remark following the theorem in § 1 the continuous functions which vanish for sufficiently large values of x are everywhere dense in $L^2_{p(x)}$. For every such function $f(x)$ the integral

$$I = \int_0^{+\infty} x^\alpha e^{-x} \left[f(x) - \sum_{k=0}^{n} c_k e^{-kx} \right]^2 dx$$

goes over into

$$I = \int_0^1 |\ln t|^\alpha \left[\varphi(t) - \sum_{k=0}^{n} c_k t^k \right]^2 dt$$

by means of the substitution [5] $x = -\ln t$; here $\varphi(t) = f[-\ln t]$ is a continuous function.[6] Since now $\varphi(t)$ can be approximated arbitrarily close (and even uniformly!) by polynomials, the integral I can be made arbitrarily small by appropriate choice of the coefficients c_k. Therefore the "polynomials" $\sum c_k e^{-kx}$

[5] Due to the special character of the function $f(x)$ the difficulties related to substitution in LEBESGUE integrals are avoided.

[6] If $f(x) = 0$ for $x \geqq A$, then $\phi(t) = 0$ for $0 \leqq t \leqq e^{-A}$. On the interval $[e^{-A}, 1]$ $\phi(t)$ is the superposition of two continuous functions.

are everywhere dense in $L^2_{p(x)}$; it now suffices to show that these "polynomials" can be approximated to arbitrary accuracy by means of polynomials (in the usual sense). To this end it is sufficient to show that each function e^{-mx} $(m = 0, 1, 2, \ldots)$ can be approximated by polynomials. For $m = 0$ this is trivial; for other values of m however it depends on the validity of the PARSEVAL equality

$$\int_0^{+\infty} x^\alpha e^{-x}(e^{-mx})^2\,dx = \sum_{n=0}^{\infty} c_n^2 \tag{230}$$

where

$$c_n = \int_0^{+\infty} x^\alpha e^{-x} e^{-mx} L_n^{(\alpha)}(x)\,dx.$$

Now

$$\int_0^{+\infty} x^\alpha e^{-(2m+1)x}\,dx = \frac{\Gamma(\alpha+1)}{(2m+1)^{\alpha+1}}.$$

If we put

$$u = \frac{(-1)^n}{\sqrt{n!\Gamma(\alpha+n+1)}} x^{\alpha+n}e^{-x}, \quad v = e^{-mx},$$

we find on the other hand that

$$c_n = \int_0^{+\infty} u^{(n)} v\,dx = [u^{(n-1)}v - \cdots + (-1)^{n-1}uv^{(n-1)}]_0^{+\infty} + (-1)^n \int_0^{+\infty} uv^{(n)}\,dx.$$

The term outside the integral vanishes, and therefore

$$c_n = \frac{(-m)^n}{\sqrt{n!\Gamma(\alpha+n+1)}} \int_0^{+\infty} x^{\alpha+n} e^{-(m+1)x}\,dx = \frac{(-m)^n}{(m+1)^{\alpha+n+1}} \sqrt{\frac{\Gamma(\alpha+n+1)}{n!}}.$$

The equality (230) which is to be proved now takes the form

$$\frac{\Gamma(\alpha+1)}{(2m+1)^{\alpha+1}} = \sum_{n=0}^{\infty} \frac{\Gamma(\alpha+n+1)}{n!} \frac{m^{2n}}{(m+1)^{2\alpha+2n+2}}. \tag{231}$$

If we further put

$$\frac{m}{m+1} = x,$$

equation (231) becomes

$$\frac{\Gamma(\alpha+1)}{(1-x^2)^{\alpha+1}} = \sum_{n=0}^{\infty} \frac{\Gamma(\alpha+n+1)}{n!} x^{2n}$$

or

$$\frac{1}{(1-x^2)^{\alpha+1}} = \sum_{n=0}^{\infty} \frac{(\alpha+n)(\alpha+n-1)\ldots(\alpha+1)}{n!} x^{2n}.$$

This is nothing but NEWTON's binomial formula. The theorem has now been proved. In particular it follows as a corollary that the orthonormal system $\{\hat{L}_n^{(\alpha)}(x)\}$ is closed in the space $L^2_{p(x)}$ with $p(x) = x^\alpha e^{-x}$.

§4. The HERMITE Polynomials

The HERMITE[7] *polynomials* form an orthogonal system with the weight function

$$p(x) = e^{-x^2} \qquad (-\infty < x < +\infty). \quad (232)$$

Theorem 1. *The* HERMITE *polynomials can be represented by the formula*

$$H_n(x) = e^{x^2} \frac{d^n(e^{-x^2})}{dx^n}. \quad (233)$$

Proof. If we put

$$u = e^{-x^2},$$

then

$$u' = -2xe^{-x^2}, \quad u'' = (4x^2 - 2)e^{-x^2}, \quad u''' = (-8x^3 + 12x)e^{-x^2}$$

and in general

$$u^{(n)} = H_n(x)e^{-x^2},$$

where $H_n(x)$ is a polynomial of degree n with leading coefficient $(-2)^n$, as one can easily prove by complete induction. The function defined by formula (233) is hence indeed a polynomial of degree n.

Now

$$u^{(n)}(-\infty) = u^{(n)}(+\infty) = 0;$$

therefore from the formula

$$\int_{-\infty}^{+\infty} u^{(n)}v\,dx = [u^{(n-1)}v - \cdots + (-1)^{(n-1)}uv^{(n-1)}]_{-\infty}^{+\infty} + (-1)^n \int_{-\infty}^{+\infty} uv^{(n)}\,dx$$

it follows that for any arbitrary polynomial $v(x)$

$$\int_{-\infty}^{+\infty} e^{-x^2} H_n(x)v(x)\,dx = (-1)^n \int_{-\infty}^{+\infty} e^{-x^2} v^{(n)}(x)\,dx. \quad (234)$$

If in particular the degree of $v(x)$ is less than n, then the last integral vanishes, and it follows that the polynomials (233) do indeed form an orthogonal system with weight e^{-x^2}.

[7] HERMITE [1].

If we substitute $v(x) = H_n(x)$ in (234) and take into account the fact that the leading coefficient of $H_n(x)$ is $(-2)^n$, then we find that

$$\int_{-\infty}^{+\infty} e^{-x^2} H_n^2(x)\,dx = 2^n n! \int_{-\infty}^{+\infty} e^{-x^2}\,dx = 2^n n!\sqrt{\pi}. \tag{235}$$

Hence

$$\left.\begin{aligned} \tilde{H}_n(x) &= \left(-\frac{1}{2}\right)^n H_n(x), \\ \hat{H}_n(x) &= \frac{(-1)^n}{\sqrt{2^n n!\sqrt{\pi}}} H_n(x). \end{aligned}\right\} \tag{236}$$

The recursion formula for the HERMITE polynomials reads

$$\tilde{H}_{n+2}(x) = x\tilde{H}_{n+1}(x) - \frac{n+1}{2}\tilde{H}_n(x).$$

For with the help of (233) it can easily be shown by complete induction that the only powers of x occurring as summands in $H_n(x)$ are those which are odd if n is odd and even if n is even; from this it clearly follows that $\alpha_{n+2} = 0$. From (73)

$$\lambda_{n+1} = \frac{\int_{-\infty}^{+\infty} e^{-x^2} \tilde{H}_{n+1}^2(x)\,dx}{\int_{-\infty}^{+\infty} e^{-x^2} \tilde{H}_n^2(x)\,dx},$$

and according to (235) and (236) therefore $\lambda_{n+1} = \frac{n+1}{2}$.

The *generating function* for the HERMITE polynomials is

$$H(t, x) = e^{-2tx-t^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n. \tag{237}$$

For the TAYLOR series expansion of $\varphi(z) = e^{-z^2}$ gives

$$\varphi(x+t) = \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(x)}{n!} t^n.$$

Since $\varphi^{(n)}(x) = e^{-x^2}H_n(x)$, (237) follows.

It is also easy to derive the differential equation satisfied by $y = H_n(x)$. For if $u = e^{-x^2}$, then $u' = -2xu$. Forming the derivative of order $n+1$, we find

$$u^{(n+2)} + 2xu^{(n+1)} + 2(n+1)u^{(n)} = 0.$$

Since $u^{(n)} = e^{-x^2}y$, we therefore have

$$y'' - 2xy' + 2ny = 0.$$

Finally we show that the system of HERMITE polynomials is complete:

Theorem 2 (V. A. STEKLOV [3]). *In the space $L^2_{p(x)}$ of the* HERMITE *weight*

$$p(x) = e^{-x^2} \qquad (-\infty < x < +\infty)$$

the set of all polynomials is everywhere dense.

Proof. The continuous functions which for sufficiently large values of $|x|$ are equal to zero lie everywhere dense in $L^2_{p(x)}$; it therefore suffices to show that these functions can be approximated by polynomials to arbitrary accuracy.

Without loss of generality we may moreover stipulate that the functions under consideration vanish on small intervals of the form $(-a, +a)$, since all such functions are also dense in $L^2_{p(x)}$.

Suppose therefore that $f(x)$ is a continuous function which is different from zero for $a < |x| < A$. We first assume that it is even. Then

$$I = \int_{-\infty}^{+\infty} e^{-x^2}\left[f(x) - \sum_{k=0}^{n} c_k x^{2k}\right]^2 dx = 2\int_{0}^{+\infty} e^{-x^2}\left[f(x) - \sum_{k=0}^{n} c_k x^{2k}\right]^2 dx.$$

The substitution $x^2 = z$ gives

$$I = \int_{0}^{+\infty} \frac{e^{-z}}{\sqrt{z}}\left[\varphi(z) - \sum_{k=0}^{n} c_k z^k\right]^2 dz.$$

The function $\varphi(z) = f(\sqrt{z})$ is continuous and bounded; it therefore belongs to the space $L^2_{p(z)}$ of the LAGUERRE weight $p(z) = z^{-\frac{1}{2}}e^{-z}$. In this space the polynomials are everywhere dense, so that by appropriate choice of the coefficients c_k the integral I can be made arbitrarily small.

If the function $f(x)$ is odd, then this same substitution gives

$$I = \int_{-\infty}^{+\infty} e^{-x^2}\left[f(x) - \sum_{k=0}^{n} c_k x^{2k+1}\right]^2 dx = \int_{0}^{+\infty} \sqrt{z}\, e^{-z}\left[\varphi(z) - \sum_{k=0}^{n} c_k z^k\right]^2 dz,$$

where $\varphi(z) = \dfrac{f(\sqrt{z})}{\sqrt{z}}$ is continuous and bounded; since now the polynomials are everywhere dense in the space $L^2_{p(z)}$ with $p(z) = z^{+\frac{1}{2}}e^{-z}$, we can again make I arbitrarily small. To complete the proof we have only to remark that every function $f(x)$ can be represented in the form

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}$$

as the sum of an even and an odd function.

§5. The Moment Problem for Infinite Intervals

The moment problem for infinite intervals is essentially more difficult than for finite intervals. While for finite intervals an integral weight of bounded variation—as we have seen above—is "almost" uniquely determined by its moments (since its kernel is uniquely determined), for infinite intervals this uniqueness is lost. For example, as we have seen in § 1

$$\int_0^{+\infty} x^{-\ln x} \sin(2\pi \ln x) x^n dx = 0 \qquad (n = 0, 1, 2, \ldots);$$

the two non-negative differential weights

$$p_1(x) = x^{-\ln x}, \quad p_2(x) = x^{-\ln x}[1 - \sin(2\pi \ln x)]$$

therefore have the same moments on the interval $[0, +\infty)$.

If we now set

$$g_i(x) = \int_0^x p_i(t)\, dt \qquad (i = 1, 2),$$

we obtain two distinct, strictly increasing, continuous integral weights which have identical moments.

We are thus faced with the problem of sorting out those moment sequences which determine their weight functions uniquely ("uniquely" in the sense previously agreed upon).

We will here leave this general problem aside and seek only condition for the special case in which a sequence of numbers $\{\mu_n\}$ constitutes a moment sequence on the interval $(-\infty, +\infty)$.

We further restrict our considerations to increasing integral weights $g(x)$ with an infinite number of points of increase. We interpret the STIELTJES integral

$$\int_{-\infty}^{+\infty} x^n dg(x), \tag{238}$$

with which we are concerned as an improper integral, i.e., as

$$\lim \int_A^B x^n dg(x) \qquad (A \to -\infty, \quad B \to +\infty).$$

Since

$$\int_A^B dg(x) = g(B) - g(A),$$

the integral (238) exists for $n = 0$ only if the weight $g(x)$ is bounded. We assume this condition to be satisfied in the sequel.

Lemma 1. *A polynomial $P(x)$ with real coefficients which is non-negative for all real values of x can be represented as a sum of squares of two polynomials with real coefficients.*

Proof. The real roots of $P(x)$ (insofar as there are any) are of even multiplicity. Denoting these roots by $c_1, c_2, \ldots, c_m$, we have

$$P(x) = Q(x)R^2(x),$$

where $R(x) = (x - c_1)^{\alpha_1}(x - c_2)^{\alpha_2} \ldots (x - c_m)^{\alpha_m}$, and $Q(x)$ has only complex conjugate roots. Two complex conjugate roots here always have the same multiplicity, and hence

$$Q(x) = A^2 \prod_{k=1}^{s} [(x - a_k - b_k i)(x - a_k + b_k i)]^{\sigma_k} = A^2 \prod_{k=1}^{s} [(x - a_k)^2 + b_k^2]^{\sigma_k}.$$

$Q(x)$ is thus a product of a certain number of factors each of which is the sum of squares of two polynomials. From the identity

$$(A^2 + B^2)(C^2 + D^2) = (AC - BD)^2 + (AD + BC)^2$$

such a product is itself the sum of squares of two polynomials. The remainder of the proof is obvious.

Theorem 1. *A sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$ if and only if all the determinants* [8]

$$\Delta_n = \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \ldots & \ldots & \ldots & \ldots \\ \mu_n & \mu_{n+1} & \ldots & \mu_{2n} \end{vmatrix} \qquad (n = 0, 1, 2, \ldots)$$

are strictly positive.

Proof. We define a functional $\Phi[P(x)]$ on the set of all polynomials such that for

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n$$

the functional

$$\Phi[P(x)] = a_0\mu_0 + a_1\mu_1 + \cdots + a_n\mu_n.$$

For this functional clearly

$$\Phi[P_1(x) + P_2(x)] = \Phi[P_1(x)] + \Phi[P_2(x)],$$
$$\Phi[kP(x)] = k\Phi[P(x)].$$

[8] We define $\Delta_0 = \mu_0$.

Now let

$$\psi_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{n-1} & 1 \\ \mu_1 & \mu_2 & \dots & \mu_n & x \\ \dots & \dots & \dots & \dots & \dots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n-1} & x^n \end{vmatrix} \quad (n = 1, 2, \dots). \tag{239}$$

In addition, let us define $\psi_0(x) = 1$. Then

$$x^k\psi_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{n-1} & x^k \\ \mu_1 & \mu_2 & \dots & \mu_n & x^{k+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n-1} & x^{k+n} \end{vmatrix},$$

and

$$\Phi[x^k\psi_n(x)] = \begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{n-1} & \Phi[x^k] \\ \mu_1 & \mu_2 & \dots & \mu_n & \Phi[x^{k+1}] \\ \dots & \dots & \dots & \dots & \dots \\ \mu_n & \mu_{n+1} & \dots & \mu_{2n-1} & \Phi[x^{k+n}] \end{vmatrix} = \begin{vmatrix} \mu_0 & \dots & \mu_{n-1} & \mu_k \\ \mu_1 & \dots & \mu_n & \mu_{k+1} \\ \dots & \dots & \dots & \dots \\ \mu_n & \dots & \mu_{2n-1} & \mu_{k+n} \end{vmatrix}.$$

Therefore

$$\Phi[\psi_n(x)] = \Phi[x\psi_n(x)] = \dots = \Phi[x^{n-1}\psi_n(x)] = 0, \tag{240}$$

$$\Phi[x^n\psi_n(x)] = \Delta_n. \tag{241}$$

From (240) it follows that for every polynomial $R(x) \in H_{n-1}$

$$\Phi[R(x)\psi_n(x)] = 0. \tag{242}$$

If on the other hand we expand $\psi_n(x)$ in terms of the elements of the last column we find

$$\psi_n(x) = \Delta_{n-1}x^n + R(x),$$

where the degree of $R(x)$ is less than n. Hence

$$\Phi[\psi_n^2(x)] = \Delta_{n-1}\Delta_n.$$

Now let us assume that the sequence $\{\mu_n\}$ is strictly positive definite. Then first of all $\Delta_0 = \mu_0 > 0$, since $\mu_0 = \Phi[1]$ and the "polynomial" 1 is non-negative and not identically zero. But then the polynomial $\psi_1(x)$ is also not identically zero (since it is $\Delta_0 x +$ const) and

$$\Delta_0\Delta_1 = \Phi[\psi_1^2(x)] > 0,$$

whence it follows that $\Delta_1 > 0$. Similarly $\psi_2(x)$ is not identically zero and

$$\Delta_1\Delta_2 = \Phi[\psi_2^2(x)] > 0,$$

and hence $\Delta_2 > 0$. Now assume that we have shown that $\Delta_{n-1} > 0$. Then $\psi_n(x)$ is not identically zero and

$$\Delta_{n-1}\Delta_n = \Phi[\psi_n^2(x)] > 0,$$

therefore $\Delta_n > 0$. Hence for a strictly positive definite sequence $\{\mu_n\}$ all the determinants Δ_n are positive.

Conversely, suppose $\Delta_n > 0$ for $n = 0, 1, 2, \ldots$. We then take any polynomial $P(x)$ which does not vanish identically. If it is of degree n, then it may be represented in the form

$$P(x) = A_0\psi_0(x) + A_1\psi_1(x) + \cdots + A_n\psi_n(x) \qquad (A_n \neq 0).$$

Then

$$\Phi[P^2(x)] = \Phi\left[\left\{\sum_{k=0}^{n} A_k\psi_k(x)\right\}^2\right] = \sum_{i,k} A_iA_k\Phi[\psi_i(x)\psi_k(x)].$$

But from (242)

$$\Phi[\psi_i(x)\psi_k(x)] = 0 \qquad (i \neq k),$$

hence [9]

$$\Phi[P^2(x)] = \sum_{k=0}^{n} A_k^2\Phi[\psi_k^2(x)] = \sum_{k=0}^{n} A_k^2\Delta_{k-1}\Delta_k > 0.$$

If now $P(x) \not\equiv 0$ is a non-negative polynomial it can according to Lemma 1 be represented in the form

$$P(x) = P_1^2(x) + P_2^2(x);$$

from what has just been shown therefore

$$\Phi[P(x)] = \Phi[P_1^2(x)] + \Phi[P_2^2(x)] > 0,$$

which implies that the sequence $\{\mu_n\}$ is strictly positive definite. This completes the proof of the theorem.

Theorem 2. *If $g(x)$ is a bounded, increasing function with an infinite number of points of increase and if all the integrals*

$$\mu_n = \int_{-\infty}^{+\infty} x^n dg(x) \qquad (n = 0, 1, 2, \ldots), \quad (243)$$

exist, then the sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$.

[9] We define $\Delta_{-1} = 1$.

Proof. If the polynomial $P(x)$ is non-negative and $P(x) \not\equiv 0$, then

$$\Phi[P(x)] = \int_{-\infty}^{+\infty} P(x)dg(x) \geqq \int_{A}^{B} P(x)dg(x),$$

where A and B are two arbitrary finite numbers. If we choose these in such a way that the number of points of increase of $g(x)$ in $[A, B]$ is greater than the degree of $P(x)$, then according to the footnote accompanying Lemma 2 of § 4, Chapter VII we obtain

$$\int_{A}^{B} P(x)dg(x) > 0,$$

from which $\Phi[P(x)] > 0$, and thus the theorem follows.

The converse is also true:

Theorem 3. (H. HAMBURGER [1]). *If the sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$, then there exists a bounded, increasing function $g(x)$ with an infinite number of points of increase which for all n satisfies the equation*

$$\mu_n = \int_{-\infty}^{+\infty} x^n dg(x) \qquad (n = 0, 1, 2, \ldots).$$

The proof of this theorem is somewhat involved. We therefore first establish several lemmas.

Lemma 2. *If the sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$, then all the roots of the polynomial* (239) *are real and simple.*

Proof. First of all, $\psi_n(x)$ has real roots of odd multiplicity. For if not, then $\psi_n(x)$ would be non-negative and in addition $\psi(x) \not\equiv 0$; but since

$$\Phi[\psi_n(x)] = 0$$

this would contradict the strict positive definiteness of $\{\mu_n\}$.

Now let $\xi_1, \xi_2, \ldots, \xi_s$ be *all* roots of $\psi_n(x)$ of odd multiplicity. If $s < n$, then we can put

$$R(x) = (x - \xi_1)(x - \xi_2) \ldots (x - \xi_s)$$

and thus obtain [cf. (242)]

$$\Phi[R(x)\psi_n(x)] = 0,$$

which likewise contradicts the strict positive definiteness of $\{\mu_n\}$, since $R(x)\psi_n(x)$ is a non-negative polynomial. Therefore $s = n$, and the lemma is proved.

Lemma 3. *Under the same conditions as in Lemma 2 the equation*

$$\Phi[P(x)] = \sum_{k=1}^{n} A_k^{(n)} P(\xi_k^{(n)}), \tag{244}$$

holds for any arbitrary polynomial $P(x) \in H_{2n-1}$, *where the arguments are the roots of* $\psi_n(x)$ *and*

$$A_k^{(n)} = \Phi\left[\frac{\psi_n(x)}{\psi_n'(\xi_k^{(n)})(x - \xi_k^{(n)})}\right].$$

Proof. We divide $P(x)$ by $\psi_n(x)$ and obtain

$$P(x) = R(x)\psi_n(x) + \rho(x),$$

where the degrees of $R(x)$ and $\rho(x)$ are less than n. From (242) therefore

$$\Phi[P(x)] = \Phi[\rho(x)].$$

But

$$\rho(x) = \sum_{k=1}^{n} \frac{\psi_n(x)}{\psi_n'(\xi_k^{(n)})(x - \xi_k^{(n)})}\, \rho(\xi_k^{(n)}),$$

since both sides of this equation are polynomials of H_{n-1} which coincide at the n points $\xi_k^{(n)}$; we have now only to note that $\rho(\xi_k^{(n)}) = P(\xi_k^{(n)})$.

Lemma 4. *Under the same conditions as in Lemma 2 the coefficients* $A_k^{(n)}$ *are positive.*

Proof. Let i be some one of the numbers $1, 2, \ldots, n$. We define

$$P(x) = \left[\frac{\psi_n(x)}{x - \xi_i^{(n)}}\right]^2.$$

This is a polynomial of degree $2n - 2$. We substitute it into (244), noting that for $k \neq i$

$$P(\xi_k^{(n)}) = 0$$

and $P(\xi_i^{(n)}) = [\psi_n'(\xi_i^{(n)})]^2$; this gives

$$\Phi[P(x)] = A_i^{(n)}[\psi_n'(\xi_i^{(n)})]^2,$$

whence it follows that $A_i^{(n)} > 0$.

We now proceed to the proof of the HAMBURGER theorem.

Proof of Theorem 3. We order the roots $\xi_k^{(n)}$ of the polynomial $\psi_n(x)$ in an increasing sequence

$$\xi_1^{(n)} < \xi_2^{(n)} < \cdots < \xi_n^{(n)}$$

and introduce a step function $g_n(x)$:

$$\begin{aligned}
g_n(x) &= 0 && \text{for } -\infty < x \leqq \xi_1^{(n)},\\
g_n(x) &= A_1^{(n)} && \text{for } \xi_1^{(n)} < x \leqq \xi_2^{(n)},\\
g_n(x) &= A_1^{(n)} + A_2^{(n)} && \text{for } \xi_2^{(n)} < x \leqq \xi_3^{(n)},\\
&\cdots\cdots\cdots

We fix i. Then for $n_m > i$

$$\mu_i = \int_{-\infty}^{+\infty} x^i dg_{n_m}(x).$$

Now let $A < 0 < B$. Then

$$\left| \mu_i - \int_A^B x^i dg_{n_m}(x) \right| \leqq \int_{-\infty}^{A} |x^i| dg_{n_m}(x) + \int_B^{+\infty} x^i dg_{n_m}(x).$$

Now let $2r$ be some even number which is greater than i. Then

$$\int_{-\infty}^{A} |x|^i dg_{n_m}(x) \leqq \frac{1}{|A|^{2r-i}} \int_{-\infty}^{+\infty} x^{2r} dg_{n_m}(x),$$

$$\int_B^{+\infty} x^i dg_{n_m}(x) \leqq \frac{1}{B^{2r-i}} \int_{-\infty}^{+\infty} x^{2r} dg_{n_m}(x),$$

hence with $M = \min \{|A|, |B|\}$:

$$\int_{-\infty}^{A} |x|^i dg_{n_m}(x) + \int_B^{+\infty} x^i dg_{n_m}(x) \leqq \frac{1}{|M|^{2r-i}} \int_{-\infty}^{+\infty} x^{2r} dg_{n_m}(x).$$

If n_m is here greater than $2r$, then the last integral is equal to μ_{2r}. For such values n_m we therefore have

$$\left| \mu_i - \int_A^B x^i dg_{n_m}(x) \right| \leqq \frac{\mu_{2r}}{M^{2r-i}}.$$

For the finite interval $[A, B]$ we may use HELLY's theorem on taking the limit under the STIELTJES integral sign; this gives

$$\left| \mu_i - \int_A^B x^i dg(x) \right| \leqq \frac{\mu_{2r}}{M^{2r-i}}.$$

As $A \to -\infty$ and $B \to +\infty$, $M \to +\infty$, so that we finally obtain

$$\mu_i = \int_{-\infty}^{+\infty} x^i dg(x)$$

It remains only to show that $g(x)$ has infinitely many points of increase. If $g(x)$ had only finitely many, then for the non-negative polynomial $P(x)$ having these points as roots the equation

$$\Phi[P(x)] = \int_{-\infty}^{+\infty} P(x) dg(x) = 0$$

would hold, but this contradicts the hypothesis that the sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$. The HAMBURGER theorem is herewith completely proved.

A simple combination of the results already obtained in this section leads to [11]

Theorem 4. *A necessary and sufficient condition that corresponding to a given sequence there exist a bounded, increasing function $g(x)$ with infinitely many points of increase such that*

$$\int_{-\infty}^{+\infty} x^n dg(x) = \mu_n \qquad (n = 0, 1, 2, \ldots)$$

is that the numbers μ_k satisfy the inequalities

$$\begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \ldots & \ldots & \ldots & \ldots \\ \mu_n & \mu_{n+1} & \ldots & \mu_{2n} \end{vmatrix} > 0. \qquad (n = 0, 1, 2, \ldots).$$

The reader will find further details in the monograph by N. I. AKHIEZER and M. G. KREIN [2]

§6. The FAVARD Theorem

The HAMBURGER theorem assists us in obtaining the

Theorem.[12] *Let $\{\omega_n(x)\}$ for $n = 0, 1, 2, \ldots$ be a polynomial system in which $\omega_n(x)$ is a polynomial of degree n with leading coefficient equal to one. If a recursion formula of the form*

$$\omega_{n+2}(x) = (x - \alpha_{n+2})\omega_{n+1}(x) - \lambda_{n+1}\omega_n(x) \quad (n = 0, 1, 2, \ldots), \quad (246)$$

holds wherein $\lambda_{n+1} > 0$, then there exists a bounded, increasing function $g(x)$ with infinitely many points of increase such that for $i \neq k$

$$\int_{-\infty}^{+\infty} \omega_i(x)\omega_k(x)dg(x) = 0.$$

In other words, from the validity of formula (246) together with the condition $\lambda_{n+1} > 0$ it follows that $\{\omega_n(x)\}$ is an orthogonal system with respect to a certain increasing integral weight.

[11] This is the original formulation of the HAMBURGER theorem.

[12] FAVARD [1]. This theorem was also discovered (independent of FAVARD) by the author (NATANSON) in the year 1935 and was presented by him in a seminar led by S. N. BERNSTEIN. He then did not publish the result, since the work of FAVARD had appeared in the meantime.

Proof Let

$$\omega_n(x) = x^n + \sigma_1^{(n)}x^{n-1} + \cdots + \sigma_n^{(n)}.$$

We now define a sequence $\{\mu_n\}$ by setting

$$\mu_0 = 1,$$

$$\mu_n + \sigma_1^{(n)}\mu_{n-1} + \cdots + \sigma_n^{(n)}\mu_0 = 0 \quad (n = 1, 2, \ldots). \quad (247)$$

Equations (247) define the numbers μ_n recursively.

With the help of the sequence $\{\mu_n\}$ we can produce a functional $\Phi[P(x)]$ as above if for the polynomial

$$P(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$$

we define

$$\Phi[P(x)] = a_0\mu_n + a_1\mu_{n-1} + \cdots + a_n\mu_0.$$

Then according to equation (247)

$$\Phi[\omega_n(x)] = 0 \qquad (n = 1, 2, 3, \ldots). \quad (248)$$

We now prove a still more general formula

$$\Phi[x^k\omega_{n+k}(x)] = 0 \qquad (n = 1, 2, 3, \ldots). \quad (249)$$

It is true for $k = 0$, since then it is identical with (248). Now suppose it has been proved for $k = m$. From the equation

$$\omega_{n+m+2}(x) = (x - \alpha_{n+m+2})\omega_{n+m+1}(x) - \lambda_{n+m+1}\omega_{n+m}(x)$$

it follows that

$$\Phi[x^m\omega_{n+2+m}(x)] = \Phi[x^{m+1}\omega_{n+m+1}(x)] - \alpha_{n+m+2}\Phi[x^m\omega_{n+1+m}(x)] - \lambda_{n+m+1}\Phi[x^m\omega_{n+m}(x)].$$

By hypothesis all of the summands except the first on the right-hand side are equal to zero. Hence this one is also equal to zero, i.e

$$\Phi[x^{m+1}\omega_{n+m+1}(x)] = 0,$$

wherewith formula (249) is proved. Changing notation, we can also write it in the form

$$\Phi[x^k\omega_n(x)] = 0 \quad (n = 1, 2, 3, \ldots;\ k = 0, 1, \ldots, n - 1)$$

Hence, for every polynomial $R(x) \in H_{n-1}$

$$\Phi[R(x)\omega_n(x)] = 0 \qquad (n = 1, 2, 3, \ldots). \quad (250)$$

From (246) on the other hand it follows that

$$\Phi[x^n\omega_{n+2}(x)] = \Phi[x^{n+1}\omega_{n+1}(x)] - \alpha_{n+2}\Phi[x^n\omega_{n+1}(x)] - \lambda_{n+1}\Phi[x^n\omega_n(x)],$$

and hence

$$\Phi[x^{n+1}\omega_{n+1}(x)] = \lambda_{n+1}\Phi[x^n\omega_n(x)].$$

Iterating on n, one obtains finally

$$\Phi[x^n\omega_n(x)] = \lambda_1\lambda_2 \ldots \lambda_n. \tag{251}$$

Now

$$\omega_n(x) = x^n + R(x),$$

where the degree of $R(x)$ is less than n. From (250) and (251) it hence follows that

$$\Phi[\omega_n^2(x)] = \lambda_1\lambda_2 \ldots \lambda_n$$

and from this, what is the important thing for us here,

$$\Phi[\omega_n^2(x)] > 0, \tag{252}$$

which is an estimate that is also valid for $n = 0$, since $\Phi[1] = 1$.

We now take an arbitrary polynomial $P(x)$ which does not vanish identically. If its degree is n, then

$$P(x) = c_0\omega_0(x) + c_1\omega_1(x) + \cdots + c_n\omega_n(x),$$

whence

$$\Phi[P^2(x)] = \sum_{i,k} c_i c_k \Phi[\omega_i(x)\omega_k(x)].$$

According to (250) all summands with $i \neq k$ vanish, and therefore according to (252)

$$\Phi[P^2(x)] = \sum_{k=0}^{n} c_k^2 \Phi[\omega_k^2(x)] > 0. \tag{253}$$

But now since every non-negative polynomial is a sum of squares of two polynomials, it follows from (253) that the sequence $\{\mu_n\}$ is strictly positive definite on $(-\infty, +\infty)$. There exists therefore a bounded, increasing function $g(x)$ with an infinite number of points of increase which satisfies the equation

$$\mu_n = \int_{-\infty}^{+\infty} x^n dg(x) \qquad (n = 0, 1, 2, \ldots).$$

For every polynomial $P(x)$

$$\int_{-\infty}^{+\infty} P(x)dg(x) = \Phi[P(x)].$$

From this and from (250) it follows that for every polynomial $R(x) \in H_{n-1}$

$$\int_{-\infty}^{+\infty} R(x)\omega_n(x)dg(x) = 0,$$

wherewith the theorem is proved [13]

[13] With reference to this and the preceding section cf. J. L. GERONIMUS [4, 5].

BIBLIOGRAPHY

AKHIEZER, N. I. and KREIN, M. G.

[2] *O nekotorykh voprosakh teorii momentov* (On certain questions in the theory of moments). Kharkov, GONTI Ukr. SSR, 1933.

CHRISTOFFEL, E. B.

[1] *Über die Gauss'sche Quadratur und eine Verallgemeinerung derselben* (On the Gauss quadrature and a generalization of the same). Journ. für Math., 1858.

DARBOUX, G.

[1] *Mémoire sur l'approximation des fonctions de très grands nombres* (Memoire on the approximation of functions of very large numbers). Journ. de Math., 1878.

FAVARD, J.

[1] *Sur les polynômes de Tchebicheff* (On the Tchebysheff polynomials). C. R. Acad. Sci., 1935.

FICHTENHOLZ, G. M.

[1] *Sur la notion de fermeture des systèmes de fonctions* (On the concept of closedness of systems of functions). Rend. Cir. Mat. Palermo, 1926.

FISCHER, E.

[1] *Sur la convergence en moyenne* (On convergence in mean). C. R. Acad. Sci., 1907.

FRÉCHET, M.

[1] *Sur les ensembles de fonctions et les opérations linéaires* (On sets of functions and linear operations). C. R. Acad. Sci., 1907.

GAGAEV, B. M.

[1] *O nekotorykh classakh ortogonal'nykh funktsiy* (On certain classes of orthogonal functions). IAN, math. series, 1946.

GERONIMUS, J. L.

[1] *On a problem of M. J. Shohat.* Amer. Journ. Math., 1932.

[2] *O nekotorykh ekstremal'nykh zadachakh* (On certain extremal problems). IAN, math. series, 1937.

[3] *Ob odnoy ekstremal'noy zadache Chebysheva* (On one of Tchebysheff's extremal problems). IAN, math. series, 1938.

[4] *Obobshchennye ortogonal'nye polinomy i formula Kristoffelya-Darbu* (Generalized orthogonal polynomials and the Christoffel-Darboux formula). DAN, 1940.

[5] *O nekotorykh svoystvakh obobshchennykh ortogonal'nykh polinomov* (On certain properties of generalized orthogonal polynomials). DAN, 1940.

GRAM, I. P.

[1] *Über die Entwicklung reeller Funktionen in Reihen mittels der Methode der kleinsten Quadrate* (On the expansion of real functions in series by the method of least squares). Journ. für Math., 1883.

HAMBURGER, H.

[1] *Über eine Erweiterung des Stieltjesschen Momentenproblems* (On an extension of the Stieltjes moment problem). Math. Ann., 1920.

HAUSDORFF, F.

[1] *Momentprobleme für ein endliches Intervall* (Moment problems for a finite interval). Math. Z., 1923.

HERMITE, C.

[1] *Sur un nouveau développement en série des fonctions* (On a new series expansion of functions). C. R. Acad. Sci., 1864.

KANTOROVICH, L. V.

[2] *K probleme momentov dlya konechnogo intervala* (Contribution to the moment problem for a finite interval). DAN, 1937.

KOLMOGOROFF, A. N.

[2] *O skhodimosti ryadov po ortogonal'nym polinonam* (On the convergence of series expansions in terms of orthogonal polynomials). DAN, 1934.

KORKIN, A. N. and ZOLATAREV, E. I.

[1] *Sur un certain minimum* (On a certain minimum). Collected works of E. I. Zolatarev, 1873.

KOROUS, J.

[1] *O rozvojí funkcí jedné réalué proměnné v řadu jistých ortogonálnich polynomů* (On the expansion of functions of one real variable in certain series of orthogonal polynomials). Rozpravy České Akademie, 1938.

LAGUERRE, E. N.

[1] *Sur l'intégrale* $\int_x^\infty \frac{e^x}{x}\,dx$ (On the integral $\int_x^\infty \frac{e^x}{x}\,dx$). Bull. Soc. Math. de France, 1879.

LEGENDRE, A. M.

[1] *Recherches sur l'attraction des sphéroides homogènes* (Studies of the attraction of homogeneous spheroids). Mém. math.-phs. présentés à l'Acad. Sci., 1785.

MÜNTZ, CH.

[1] *Über den Approximationssatz von Weierstrass* (On the Weierstrass approximation theorem). Schwarz-Festschrift, 1914.

NATANSON, I. P.

[1] *K voprosu o razlozhenii funktsiy po ortogonal'nym polinomam* (On the question of expanding functions in terms of orthogonal polynomials). IAN, math.-phys. series, 1933.

[2] *O skhodimosti ryadov po ortogonal'nym polinomam* (On the convergence of series expansions in terms of orthogonal polynomials). DAN, 1934.

[5] *Theory of functions of a real variable* (English translation). Vol. I, Frederick Ungar, New York, 1961.

NIKOLAEV, V. F.

[1] *K voprosu o priblizhenii nepreryvnykh funktsiy polinomami* (On the question of approximating continuous functions by means of polynomials). DAN, 1948.

PEEBLES, G.

[1] *Some generalizations of the theory of orthogonal polynomials.* Duke Math. Journal, 1940.

RADEMACHER, H.

[1] *Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen* (Several theorems on series of general orthogonal functions). Math. Ann., 1922.

RIESZ, F.

[1] *Sur les systèmes orthogonaux de fonctions* (On orthogonal systems of functions).

[2] *Sur les opérations fonctionelles linéaires* (On linear functional operations). C. R. Acad. Sci., 1909.

RODRIGUES, O.

[1] *Corresp. sur l'école polytechnique* (Correspondence to the polytechnical school). 1816.

SCHMIDT, E.

[1] *Entwicklung willkürlicher Funktionen nach vorgeschriebenen Systemen* (Expansion of arbitrary functions in terms of prescribed systems). Math Ann., 1907.

SHOHAT, J.

[1] *On a general formula in the theory of Tchebysheff polynomials and its applications.* Trans. Amer. Math. Soc., 1927.

STEKLOV, V. A.

[1] *Sur la théorie de fermeture des systèmes de fonctions orthogonales dépendant d'un nombre quelconque de variables* (On the theory of closedness of systems of orthogonal functions depending on an arbitrary number of variables). Zap. AN, 1911.

[3] *Théorème de fermeture pour les polynômes de Laplace-Hermite-Tchebychef* (Closedness theorem for the Laplace-Hermite-Tchebysheff polynomials). IAN, 1916.

[4] *Théorème de fermeture pour les polynômes de Tchebychef-Laguerre* (Closedness theorem for the Tchebysheff-Laguerre polynomials). IAN, 1916.

STIELTJES, T. J.

[1] *Quelques recherches sur la théorie des quadratures dites mécaniques* (Certain investigations of the so-called mechanical quadratures). Complete Works, 1884.

[2] *Recherches sur les fractions continues* (Investigations of continued fractions). Complete Works, 1894.

SZEGÖ, G.

[1] *Orthogonal polynomials.* Amer. Math. Soc. Colloq. Pub., 1939.

TOEPLER, A.

[1] *Bemerkenswerte Eigenschaften der periodischen Reihen* (Remarkable properties of periodic series). Wiener Akad. Anz., 1876.

INDEX

Choice
  HELLY'S principle of . 130
Completeness of the space 13
Convergence
  in itself 12
  in mean 11
  weak 15
Dense, everywhere . . 15
Determinant, GRAM 33
Differential weight . . . 127
Distance 11
Eigenfunction 20
Eigenvalue 20
Equality, PARSEVAL . . 26
Formula
  CHRISTOFFEL-DARBOUX 66
  RODRIGUES 82
Function
  generating 89
  normal 129
  LEBESGUE 72
  step 15
Functional, linear 136
Inequality
  BESSEL 26
  SCHWARZ 9
Integral
  EULER 104
  kernel of 67
  LAPLACE 92
  STIELTJES . 130
  weight 127
Kernel of the integral 67
Limit of a sequence 11
Linear functional 136
Linearly (in)dependent . 32
Localization principle 70
Metric space 11
Moment
  problem 127
  sequence 127
Norm of the function . . 10
Orthogonal 19
  polynomials 45
Orthogonal system
  closed 29
  normalized 19
Orthogonalization theorem
  E. SCHMIDT 35
Polynomial
  HERMITE 156
  JACOBI 102
  LAGUERRE 151
  LEGENDRE 82
  orthogonal 45
  TCHEBYSHEFF 114
  ultraspherical 102
Positive definite 142
Principle of choice, HELLY'S . 130
Series, FOURIER 24
Space
  $L^2_{p(x)}$ 11
  metric 11
Step function 15
System
  fundamental 39
  orthonormal 19
  RADEMACHER 22
  STURM-LIOUVILLE 20
System of functions, complete . . 30
Ultraspherical polynomial 114
Weight 9
  differential 127
  function 9
  HERMITE 150
  integral 127
  LAGUERRE 150

INDEX OF SYMBOLS

$B(p, q)$ 104
C 15
$\|f\|$ 10
$H_n(x)$ 156
$J_n^{(\alpha,\beta)}(x)$ 102
L, L^2 9
$L([a, b]), L^2([a, b])$ 9
$L_{p(x)}, L^2_{p(x)}$ 9
$L_{p(x)}([a, b)], L^2_{p(x)}[(a, b])$ . . 9
$L_n(x)$ 72, 151
$L_n^{(\alpha)}(x)$ 153
M 15
P 15
$P_n(x)$ 86
$P_n^{(\alpha,\beta)}(x)$ 107
S 15
$S_n[f] = S_n[f; x]$ 27
T 15
$U_n(x)$ 114
$W_n(x)$ 122
$X_n(x)$ 82
Δ_n 33, 45
$\Delta^n \mu_k$ 130
μ_n 45
$\Phi(f)$ 136
$\hat{T}$ 73

continued from front flap

consideration, however, are closely interwoven, and it is this inner connection that makes Constructive Function Theory, even apart from its practical importance, one of the most fascinating branches of all mathematics.

PROFESSOR NATANSON, a master of readable style, has preferred to go into detail even at the sacrifice of space rather than jeopardize clarity by too scant a presentation. The able translation provides English-reading students with the opportunity to familiarize themselves with the work of a leading Russian mathematician.

Books on Mathematics

AKHIEZER, N. I. and GLAZMAN, I. M. *Theory of Linear Operators in Hilbert Space*. Vol. I, Translated by Merlynd Nestell. *$6.25*

—Vol. II *$7.50*

BAKER, HENRY F. *Principles of Geometry:* Vol. I: *Foundations*. *$5.50*

—Vol. II: *Plane Geometry, Conics, Circles, Non-Euclidean Geometry*. *$6.50*

—Vol. III: *Solid Geometry, Quadrics, Cubic Curves in Space, Cubic Surfaces*. *$6.50*

—Vol. IV: *Higher Geometry*. Being illustrations of the utility of the consideration of higher space, especially of four and five dimensions. *$7.50*

—Vol. V: *Analytical Principles of the Theory of Curves*. *$6.50*

—Vol. VI: *Algebraic Surfaces*. *$7.50*

DEAUX, ROLAND. *Introduction to the Geometry of Complex Numbers*. Translated by Howard Eves. *$6.50*

FRANZ, WOLFGANG. *General Topology*. Vol. I. Translated by Leo F. Boron and others. *$7.50*

GOODSTEIN, R. L. *Mathematical Logic*. *$3.75*

—*Constructive Formalism:* Essays on the Foundations of Mathematics. *$3.00*

GOODSTEIN, R. L. and PRIMROSE, E. J. F. *Axiomatic Projective Geometry*. *$3.25*

HERLAND, LEO. *Dictionary of Mathematical Sciences*. Vol. I—German-English. Enlarged revised edition. *$8.50*

—Vol. II—English-German. *$6.50*

LIUSTERNIK, L., and SOBOLEV, V. *Elements of Functional Analysis*. Translated by A. E. Labarre, Jr., Herbert Izbicki, and Ward Crowley. *$8.50*

NATANSON, I. P. *Constructive Function Theory*. Vol. I: *Uniform Approximation*. Translated by Alexis N. Obolensky. *$8.00*

—*Constructive Function Theory*. Vol. II. *Approximation in Mean*. Translated by John R. Schulenberger. *$8.00*

—*Constructive Function Theory*. Vol. III: *Interpolation and Approximation Quadratures*. Translated by John R. Schulenberger. *$8.00*

—*Theory of Functions of a Real Variable*. Vol. I (chapters I to IX). Translated and edited by Leo F. Boron and Edwin Hewitt. *$8.50*

—*Theory of Functions of a Real Variable*. Vol. II (chapters X to XVII with appendices). Translated and edited by Leo F. Boron. *$8.50*

RIESZ, FRIGYES, and SZ.-NAGY, BÉLA. *Functional Analysis*. Translated by Leo F. Boron. *$10.50*

—*Appendix to Functional Analysis:* Extensions of Linear Transformations in Hilbert Space Which Extend Beyond This Space, by Béla Sz.-Nagy. *$2.50*

VAN DER WAERDEN, B. L. *Modern Algebra*. Revised Edition.

—Vol. I—translated by Fred Blum. *$6.00*

—Vol. II—translated by Theodore J. Benac. *$5.50*

WAISMANN, FRIEDRICH, *Introduction to Mathematical Thinking*. Translated by Theodore J. Benac. *$5.00*